

**CARACTERIZACIÓN DE LAS VARIABLES EXPLICATIVAS DEL CONSUMO  
ENERGÉTICO EN LAS TRECE SUBREGIONES DEL DEPARTAMENTO DE  
NARIÑO**

**JAIRO ANDRÉS CUNDAR CHECA  
DARÍO FERNANDO FAJARDO FAJARDO**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA  
CONVENIO UNIVERSIDAD DE NARIÑO  
FACULTAD DE INGENIERÍA INDUSTRIAL  
MAESTRÍA EN INVESTIGACIÓN DE OPERACIONES Y ESTADÍSTICA  
SAN JUAN DE PASTO**

**2019**

**CARACTERIZACIÓN DE LAS VARIABLES EXPLICATIVAS DEL CONSUMO  
ENERGÉTICO EN LAS TRECE SUBREGIONES DEL DEPARTAMENTO DE  
NARIÑO**

**JAIRO ANDRÉS CUNDAR CHECA  
DARÍO FERNANDO FAJARDO FAJARDO**

**Trabajo de grado presentado como requisito para optar al título de  
Magister en Investigación de Operaciones y Estadística**

**Director:**

**Mg. JOHANN ALEXIS OSPINA GALINDEZ**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA  
CONVENIO UNIVERSIDAD DE NARIÑO  
FACULTAD DE INGENIERÍA INDUSTRIAL  
MAESTRÍA EN INVESTIGACIÓN DE OPERACIONES Y ESTADÍSTICA  
SAN JUAN DE PASTO**

**2019**

### **Nota de Responsabilidad**

Las ideas y conclusiones aportadas en este Trabajo de Grado son Responsabilidad de los autores.

Artículo 1 del Acuerdo No. 324 de octubre 11 de 1966, emanado por el Honorable Concejo Directivo de la Universidad de Nariño.

Nota de Aceptación:

---

---

---

---

---

---

Firma del Presidente del Jurado

---

Firma del Jurado

---

Firma del Jurado

San Juan de Pasto, septiembre de 2019

### **Agradecimientos**

A la Universidad Tecnológica de Pereira por darnos la oportunidad de recibir la formación específica en estadística e investigación de operaciones

A la Universidad de Nariño por abrir sus puertas para que se implementen programas acreditados de alta calidad

Al MSc Hernán Abdón García, por su colaboración durante todo el desarrollo de la maestría

Al PhD. José Soto Mejía, por su colaboración en el desarrollo de las actividades y coordinación de la maestría.

Al MSc Johann Alexis Ospina Galindez por su asesoría y orientación durante el desarrollo de la tesis.

## Resumen

El consumo de energía es uno de los principales factores que determina el desarrollo económico de una región además que afecta de manera directa la calidad de vida y el avance hacia el progreso de una sociedad, es por eso que se ha desarrollado un estudio acerca de las principales fuentes de energía utilizadas en las diferentes regiones del departamento de Nariño con índice de ruralidad alto (mayor o igual a 40), de acuerdo a lo establecido en el Plan de Energización Rural para el departamento de Nariño y las definiciones de ruralidad establecidas por el PNUD, con esto se caracterizaron las variables que mejor explican el consumo energético dentro de esta población, analizando sus componentes principales; la caracterización en grupos de consumidores con características similares en cuanto al consumo y también las proyecciones hacia futuro y el impacto en el desarrollo de estas comunidades.

Se realiza un análisis univariado, un análisis correlacional y un modelo de regresión logística de las variables que explican el consumo de energía eléctrica en el departamento de Nariño, además se incluyeron variables de tipo ambiental como son irradiancia global horizontal e intensidad de la velocidad del viento, obtenidas de los mapas energéticos de Nariño.

Los resultados indicaron que el consumo eficiente en los hogares rurales de Nariño, puede ser modelado por cuatro variables de las condiciones habitacionales y hábitos de consumo, igualmente se pudieron obtener correlaciones significativas entre diferentes variables de distinta naturaleza.

**Palabras clave:** consumo energético, caracterización de grupos, metodología CRISP-DM, modelamiento.

### **Abstract**

Energy consumption is one of main factors that determines economic development of a region, which also directly affects life quality towards the society progress, which is why a study has been developed about main energy sources used in different subregions of Nariño with a high rurality index (greater than or equal 40), in accordance with related in Sustainable Rural Energy Plan (PERS) for the Nariño department and rurality definitions of PNUD. Variables that best explain the energy consumption were characterized, analyzing its main components; characterization in groups of consumers with similar characteristics in terms of consumption, and also. the projections towards the future and the impact on the development of these communities.

A univariate analysis, a correlational analysis and a logistic regression model of the variables that explain consumption of electrical energy in Nariño department Nariño are performed, in addition environmental variables were included such as horizontal global irradiance and wind speed intensity, obtained from the energy maps of Nariño.

The results indicated that efficient consumption in rural households in Nariño can be modeled by four variables of housing conditions and consumption habits, as well as significant correlations between different variables of different nature.

**Key words:** energy consumption, group characterization, CRISP-DM methodology, modeling.

## Contenido

	Pág.
1. Aspectos Preliminares .....	13
1.1 Introducción .....	13
1.2 Planteamiento del problema .....	13
1.3 Delimitación del Problema.....	14
1.4 Viabilidad de la Investigación.....	15
1.5 Limitación de la Investigación .....	15
1.6 Justificación.....	15
1.7 Antecedentes del Tema .....	16
1.8 Objetivos .....	19
1.8.1 Objetivo General .....	19
1.8.2 Objetivos Específicos .....	19
2. Conceptualización del problema de investigación .....	19
2.1. Población de estudio .....	19
2.1.1 Localización y división del departamento de nariño.....	19
3. Materiales y Métodos .....	26
3.1. Comprensión del Problema .....	26
3.1.1. Contexto .....	26
3.1.2. Objetivo.....	27
3.2. Comprensión de los Datos .....	27
3.2.1 Metodología de recolección de información primaria .....	28
3.3 Descripción del Diccionario de Datos Inicial .....	29



3.3.1 Limpieza de datos .....	32
3.3.2 Transformación. ....	35
3.4 Resultados .....	36
3.4.1 Análisis univariado.....	36
3.4.2 Análisis bivariado.....	55
3.4.3 Modelación del consumo .....	60
3.4.4 Análisis correlacional de las variables numéricas presentes en la base de datos inicial. .....	62
4. Conclusiones .....	79
5. Recomendaciones.....	81
Bibliografía .....	83
Anexos.....	85

**Lista de tablas****Pág.**

Tabla 1.	Características de las Subregiones Departamento de Nariño - .....	21
Tabla 2.	Variables Encuesta PERS Nariño .....	29
Tabla 3.	Descripción de variables en la investigación .....	33
Tabla 4.	Consumo de energía según características del hogar.....	55
Tabla 5.	Consumo de energía según dispositivos de comunicación. ....	57
Tabla 6.	Consumo de energía según características de conexión y electrodomésticos. ....	58
Tabla 7.	Consumo de energía según características de la forma de cocinar. ....	59
Tabla 8.	Estimación de los parámetros del modelo inicial.....	61
Tabla 9.	Estimación de los parámetros del modelo final. ....	62
Tabla10.	Correlaciones más importantes del análisis.....	80

### Lista de figuras

	<b>Pág.</b>
Figura 1. Distribución por sub regiones departamento de Nariño .....	24
Figura 2. Características consumo- Composición hogares .....	37
Figura 3. Consumo hogares-Estrato, ingresos, estado vivienda .....	38
Figura 4. Uso vivienda, material de pared, material piso, procedencia de agua .....	39
Figura 5. Alcantarillado, sanitaria, manejo de basuras, lugar de cocina .....	41
Figura 6. Número de: salas, cocinas, comedores baños .....	42
Figura 7. Número de: habitaciones, garajes, bodegas .....	44
Figura 8. Uso aparatos electricos .....	45
Figura 9. Consumo en viviendas .....	46
Figura 10. Consumo en viviendas: Lamparas ahorradoras y estufa eléctrica .....	47
Figura 11. Promedio consumo electrico y valor pagado o recordado .....	48
Figura 12. Condiciones del servivcio de energía eléctrica .....	49
Figura 13. Interrupciones del servicio, recibo energía, disposición a pagar por el servicio .....	50
Figura 14. Dispuesto a pagar por acometida, Cuanto esta dispuesto a pagar .....	51
Figura 15. Fuentes de energía .....	52
Figura 16. Consumo para cocción .....	53
Figura 17. Distribución de encuestas .....	54
Figura 18. Relación entre recurso solar ( $\text{kW/m}^2$ ) y velocidad del viento ( $\text{m/s}^2$ ) .....	63
Figura 19. Mapa del recurso solar para el Departamento de Nariño .....	64
Figura 20. Mapa del recurso eólico para el Departamento de Nariño .....	64
Figura 21. Relación entre consumo de energía en aparatos eléctricos e ingresos mensuales en los hogares .....	65

Figura 22.	Relación entre el consumo de leña e ingresos mensuales de los hogares.....	66
Figura 23.	Relación entre ingresos mensuales y uso de energía en refrigeración .....	67
Figura 24.	Relación entre el uso de lámparas ahorradoras e ingresos mensuales en el hogar ..	67
Figura 25.	Relación entre el material predominante en pisos e ingresos mensuales en el hogar .....	68
Figura 26.	Relación entre el uso de lámparas ahorradoras y uso de aparatos eléctricos .....	69
Figura 27.	Relación entre el uso de aparatos eléctricos y el material predominante en los pisos de la vivienda.....	70
Figura 28.	Relación entre el uso leña y refrigeración en el hogar .....	71
Figura 29.	Relación entre el consumo de leña vs material predominante pisos .....	72
Figura 30.	Relación Relación entre refrigeración y lámparas ahorradoras .....	72
Figura 31.	Relación entre refrigeración vs el material predominante en pisos .....	73
Figura 32.	Relación entre refrigeración vs número de comedores .....	74
Figura 33.	Relación entre refrigeración vs número de salas por hogar.....	75
Figura 34.	Lámparas ahorradoras vs material predomínate en pisos .....	76
Figura 35.	Lámparas ahorradoras vs número de baños por hogar .....	76
Figura 36.	Lámparas ahorradoras vs número de comedores por hogar .....	77
Figura 37.	Horas al día servicio de energía vs días en semana de servicio eléctrico.....	77
Figura 38.	Matriz de correlación .....	78

## **1. Aspectos Preliminares**

### **1.1 Introducción**

Actualmente existe una gran preocupación por estimar la demanda de energía eléctrica de una determinada población, por lo cual se han llevado a cabo estudios, como (Luna & Cevallos, 2015) , (Karatasoua & Santamouris, 2019) que han permitido reconocer algunos patrones de consumo energético, de acuerdo a las características socioeconómicas, ambientales y de ubicación de los consumidores.

Esta investigación tuvo como finalidad, descubrir los factores que describen el consumo energético de la población rural del departamento de Nariño, para lo cual se utilizó un estudio realizado en el año 2013 por el PERS Nariño (Plan de Energización Rural Sostenible), el cual trabaja con una muestra de 2479 registros y con 121 variables entre las que destacan las variables socioeconómicas, de ubicación, condiciones del servicio de energía eléctrica, las de consumo eléctrico, y utilización de otro tipo de fuentes de energía, las cuales fueron utilizadas en un estudio de minería de datos que permitió el descubrimiento de factores que más influyen en el consumo energético de las regiones rurales del departamento.

### **1.2 Planteamiento del problema**

Diferentes condiciones sociodemográficas, junto con la aparición de zonas pobladas alejadas de las concentraciones urbanas ha llevado a un consumo cada vez mayor de energía, que es obtenida de varias fuentes, la mayoría de ellas no renovables como, por ejemplo, la combustión de leña, la utilización de petróleo y sus derivados, carbón entre los más destacados. Estudios como (Aydin & Brounen , 2019) resaltan el problema del cambio climático y la dependencia de la energía en el sector residencial. En zonas más cercanas a las urbanizaciones se hace uso de la energía eléctrica del sistema interconectado para la mayoría de las actividades comerciales, industriales

institucionales y residenciales, sin dejar de lado la utilización de fuentes alternativas antes mencionadas para realizar las actividades cotidianas sobre todo en lo que refiere al consumo residencial.

El departamento de Nariño gracias a su condición geográfica, departamento costero, variedad de pisos térmicos, diversidad de pobladores y diferentes fuentes de producción, como agricultura ganadería y minería entre los principales, se ha dividido en trece subregiones con características particulares lo que ha facilitado su estudio en muchos aspectos no solo productivos sino también culturales. Debido a esto se ha hecho necesario varios estudios que permitan identificar cuáles son las necesidades primordiales de sus habitantes, sobre todo en las zonas rurales, esto ha permitido conocer más a fondo sus características de producción, consumo y las dificultades en cuanto al desarrollo. Uno de los estudios importantes que se ha realizado es el denominado Plan de energización rural sostenible de Nariño (PERS) el cual a través de la aplicación de encuestas y mediciones evaluó el consumo energético en las trece subregiones del departamento, recolectando información importante que permitió estimar variables de los consumidores en la zona rural y de proyectar el consumo energético teniendo en cuenta variables como el crecimiento demográfico y el desarrollo económico.

En este sentido definimos la pregunta de investigación: ¿Cómo identificar las variables que influyen en el consumo energético en las zonas rurales del departamento de Nariño?

### **1.3 Delimitación del Problema**

Esta investigación se basa en un estudio realizado por el Plan de Energización Rural Sostenible (PERS Nariño) del año 2013, en el cual se desarrolla una recolección de información primaria acerca de variables socioeconómicas y de consumo energético en las trece subregiones del

departamento de Nariño, a la base de datos obtenida en este estudio se le agregaron variables ambientales como son la irradiancia global horizontal y la velocidad del viento, ya que estas representan fuentes de energía local, que podrían facilitar una alternativa de generación de energía limpia, lo cual posibilita analizar la relación de estos factores con el consumo energético y el estilo de vida de sus habitantes.

#### **1.4 Viabilidad de la Investigación**

La investigación resulta viable ya que la universidad de Nariño tiene acceso a la base de datos de consumo energético generada por el PERS Nariño en el año 2013, dado que esta institución contribuyó a la recolección de la información y al estudio de este resultado.

#### **1.5 Limitación de la Investigación**

Gran parte de las variables de la investigación no aportan información pues contienen registros vacíos que no fueron contestados en la encuesta, además existen diferencias en cuanto al tipo de variables numéricas, categóricas, continuas o de caracteres.

#### **1.6 Justificación**

Actualmente se hace uso extensivo de combustibles fósiles para suplir el servicio de energía en zonas aisladas geográficamente y distantes del sistema interconectado nacional, además el costo del Kilovatio hora en estas zonas aisladas es casi el triple del precio regular según la comisión de regulación de energía y gas (CREG), situación que resulta costosa para el gobierno colombiano, existiendo particular interés en identificar las variables que explican el consumo eficiente de energía en los hogares. De acuerdo a los reportes que realiza la Unidad de Planeación Minero Energética sobre la proyección de la demanda, los bienes durables como la iluminación, el aire acondicionado y refrigeración están generando mayor consumo de energía residencial, teniendo un peso mayor sobre el consumo total y como consecuencia causa una reducción de la Intensidad

Energética en el país (UPME 2019). En estos informes periódicos se menciona que el consumo de energía eléctrica de Colombia es función directa del crecimiento económico.

Tras la revisión de literatura sobre el estudio del consumo energético se ha observado la relación existente entre el desarrollo económico y el uso eficiente de la energía, que de acuerdo al Departamento Nacional de Planeación (2017), el consumo energético eléctrico y la relación que existe con el desarrollo económico medido a través del producto interno bruto. En este estudio también se enfatiza acerca de la eficiencia de los recursos energéticos primarios como petróleo, carbón, gas biocombustibles.

En muchas ocasiones la falta de desarrollo económico y social se refleja en la ausencia de condiciones eléctricas favorables, como por ejemplo la presencia o no de interconexión eléctrica, la cual es muy escasa o nula en regiones apartadas sobre todo en la costa pacífica nariñense, en donde el coste de implementación de este sistema resulta muy elevado, es por eso que en estas zonas la mayoría de las familias utilizan como fuentes de energía, diésel y gasolina que contaminan el medio ambiente.

### **1.7 Antecedentes del Tema**

El consumo de electricidad es una variable de amplio interés para organismos de gobierno que planifican programas regionales y nacionales eficientes en búsqueda de garantizar un servicio confiable para una población. Esta demanda de energía se encuentra ligada de manera muy fuerte a factores socioeconómicos, ambientales y culturales de una región, su estudio tiene particular interés tanto para las proyecciones de los prestadores del servicio, así como para los interesados en la ampliación del mercado energético o la introducción de nuevas fuentes locales de energía.

Particularmente se ha analizado el consumo de electricidad en edificios residenciales, mostrando que variables asociadas al uso de electrodomésticos y sociodemográficas tienen un



poder explicativo de 21% y 34% respectivamente (Huebner, Shipworth, Hamilton, Chalabi y Oreszczyn 2016). Las dimensiones de la vivienda junto a las variables asociadas a electrodomésticos resultaron ser las variables más fuertemente predictoras del consumo de energía.

Por otra parte, (Jones, Fuertes y Lomas 2014) investigaron si los factores de vivienda, socioeconómicos y electrodomésticos tienen o no significancia sobre el consumo de energía. Los resultados muestran que no menos de 62 factores tienen efecto potencial sobre el consumo de energía doméstico, 13 de ellos de tipo socioeconómico, 12 respecto a la vivienda, y 37 asociados al uso de electrodomésticos. Actualmente (Karatasoua y Santamouris 2019) encontraron que el tamaño del hogar está asociado positivamente con el consumo de energía residencial con un efecto directo. El estrato socioeconómico tiene un fuerte impacto en este consumo, medianamente el tamaño de la unidad de vivienda y la cantidad de electrodomésticos que posee la vivienda.

Son muchos los esfuerzos por predecir de la mejor manera la demanda de energía residencial en una población para la toma de decisiones a corto y largo plazo, se ha realizado una revisión de las experiencias de esta actividad de interés (Swan y Ugursal 2009), donde se abordan dos enfoques, El denominado Top-Down que trata al sector residencial como un sumidero de energía y no se ocupa de los usos finales individuales usando valores históricos de la energía y la realimentación del consumo de energía del conjunto de viviendas en función de variables de alto nivel tales como los indicadores macroeconómicos (por ejemplo, el producto interno bruto, el desempleo y la inflación), el precio de la energía y el clima general. El segundo enfoque es el Botton-Up que extrapola el consumo de energía estimado de un conjunto representativo de viviendas individuales a niveles regionales y nacionales, donde se aplican el método estadístico y el método de ingeniería en la estimación.

Otro reporte en la estimación de la demanda se desarrolla en (Rahman, Sajib, Rifat, Haider y Khan 2016) donde se describe la predicción a largo plazo en Bangladesh a partir de variables como el PIB per cápita, el precio del petróleo, el crecimiento de la población y el consumo de electricidad histórico. Luego, (Ahmad, Larijani, Emmanuel, Mannion y Javed 2017) reportan un modelo desarrollado a partir de una red neuronal previamente entrenada con resultados de muy buena precisión. Se toman 10 datos de entrada de las mediciones de variables del ambiente de las habitaciones de un edificio, y el número de ocupantes. Los conjuntos de datos son generados a partir de los datos medidos y el modelo de la RNN (Red Neuronal Recurrente) es entrenado con estos conjuntos de datos para la predicción de la demanda de energía. Como un esfuerzo semejante (Subbiah, Pal, Nordberg, Marathe y Marathe 2017) desarrollan un modelo con muy buenos resultados para una ciudad basado en tres enfoques: la actividad de las personas, el nivel de edificación y el tiempo de usos de equipos o aparatos. Los resultados se pueden extender a otras ciudades mostrando robustez.

Como un enfoque de gestión de la demanda en una microrred se han adelantado trabajos en la búsqueda de acoplar generación distribuida como es el caso (Stephens, Smith y Mahanti, 2017) donde aplica el control de modelo predictivo (MPC) para la gestión de la demanda con mejores resultados que la actividad de ajuste día a día habitual. Se optimizan efectivamente los recursos en relación a una microrred acoplando tanto la demanda de energía como la generación distribuida con resultados apreciables aplicados a una microrred. Esta aproximación robusta logra mayores ahorros de costos de electricidad y la reducción de la relación pico a promedio de la demanda en comparación con el esquema de optimización día a día habitual.

## **1.8 Objetivos**

### **1.8.1 Objetivo General**

Caracterizar las variables socioeconómicas, demográficas y de uso de energía que explican mejor el consumo de energía en el sector rural del departamento de Nariño.

### **1.8.2 Objetivos Específicos**

- ✓ Identificar los diferentes grupos de consumidores de energía de acuerdo a la información obtenida en campo para el departamento de Nariño.
- ✓ Realizar un análisis de factores respecto al consumo de energía en la zona rural del departamento de Nariño.

## **2. Conceptualización del problema de investigación**

### **2.1. Población de estudio**

#### **2.1.1 Localización y división del departamento de Nariño**

El Departamento de Nariño creado en el año 1904, se encuentra ubicado al sur occidente de Colombia, en la frontera con el Ecuador. Tiene una superficie de 33.268 km<sup>2</sup> lo que representa el 2,9 % del territorio nacional. Sus límites son: por el Norte con el Departamento del Cauca, por el Sur con la República del Ecuador, por el Este con el Departamento del Putumayo y por el Oeste con el Océano Pacífico.

Nariño está integrado por tres grandes regiones geográficas de Colombia: la Llanura del Pacífico en el sector oriental, que representa una extensión del 52% del Departamento, la Región Andina que atraviesa el Departamento por el centro de norte a sur, que representa el 40% del territorio, y la Vertiente Amazónica ubicada al sur oriente del mismo, con el 8% de la extensión territorial de Nariño.

Su posición astronómica está entre los 0° 21´ y 2° 40´ de latitud norte (desembocadura del río Iscuandé); y entre los 76° 50´ (cerro Petacas y Ánimas) y 79° 02´ (cabo Manglares) de longitud oeste, del meridiano de Greenwich.

El departamento esta ordenado administrativamente en 13 subregiones, siendo San Juan de Pasto, su capital. En total el departamento tiene 64 municipios que se ubican en las subregiones administrativas: Sanquianga, Pacifico Sur, Telembí, Pie de Monte Costero, Exprovincia de Obando, Sabana, Abades, Occidente, Cordillera, Centro, Juanambú, Rio Mayo y Guambuyaco.

Las trece subregiones fueron definidas en el plan de desarrollo departamental “Nariño Mejor” 2012 – 2015 como una manera para consolidar un desarrollo endógeno con base a las capacidades económicas, sociales, institucionales e iniciativas diferenciales en las subregiones y municipios del departamento. Aunque la interrelación ambiental y cultural ha configurado a Nariño como un departamento de Costa, Sierra y Pie de Monte Costero, las 13 subregiones, claramente identificables se describen a continuación.

**Tabla 1.**  
**Características de las Subregiones Departamento de Nariño -**

Subregión	Municipios que la conforman	Extensión		Población							Actividades económicas	Cultivos
		Área (Km <sup>2</sup> )	Porcentaje del total en el depto.	Urbana	Rural	Porcentaje del total en el depto..	Hombres	Mujeres	Población indígena	Afrocolombianos		
<b>Sanquianga</b>	El Charco, La Tola, Mosquera, Olaya Herrera y Santa Bárbara	5.844	16,81%	33.258	68.455	6,13%	52%	48%	1778	67165	Agrícola, pesquera, minera, la explotación de ganado bovino y especies menores	Plátano, coco, cacao, caña de azúcar y frutas tropicales
<b>Pacífico Sur</b>	Tumaco y Francisco Pizarro	4.734	13,61%	106.747	89.572	11,83%	50%	50%	7461	139496	Agropecuaria y pesquera(cría de camarones, los enlatados), minera, ganado bovino y especies menores	Plátano, palma africana, coco, cacao, banano, arroz, maíz y frutas tropicales
<b>Telembí</b>	Barbacoas, Roberto Payán y Magüí Payán	6.206	17,85%	19.497	55.084	4,49%	52%	48%	5200	48156	Minera, agropecuaria	Arroz, cacao, plátano y frutas tropicales
<b>Piedemonte Costero</b>	Ricaurte y Mallama	2.953	8,49%	3.624	21.804	1,53%	52%	48%	17668	171	Agrícola, explotación de ganado bovino y especies menores, actividad artesanal y minera.	Maíz, plátano, caña panelera, yuca, café, frijol, y frutas
<b>Exprovincia de Obando</b>	Ipiales, Aldana, Guachucal, Cumbal, Cuaspud, Pupiales, Puerres, Cordoba, Potosí, El Contadero, Iles,	4.894	14,07%	124.646	149.130	16,49%	50%	50%	92059	268	Agrícola, explotación de ganado bovino, porcino, caballar y especies menores, actividad comercial y artesanal	Papa, maíz, trigo, cebada, frijol, alverja

	Gualmatán y Funes											
<b>La Sabana</b>	Túquerres, Imués, Guaitarilla, Ospina y Sapuyes	643	1,85%	25.712	49.980	4,56%	50%	50%	15358	3404	Agropecuaria, explotación de ganado bovino, porcino, caballar y especies menores, actividad comercial y artesanal.	Papa, trigo, cebada, maíz, arveja, hortalizas y pastos mejorados
<b>Los Abades</b>	Samaniego, Santacruz y Providencia	1.362	3,92%	29.411	58.025	5,27%	51%	49%	10119	2313	Agropecuaria, explotación de ganado bovino y especies menores, actividad minera y artesanal.	Caña de azúcar, café, maíz, frijol, cebolla cabezona, plátano, cítricos y frutales
<b>Occidente</b>	Sandoná, Linares, Consacá y Ancuya	452	1,30%	17.298	36.539	3,24%	52%	48%	14	106	Agropecuaria, explotación de ganado bovino, porcino y especies menores, actividad artesanal y comercial.	Café, plátano, maíz, yuca, fique, caña de azúcar y frutales
<b>Cordillera</b>	Taminango, Policarpa, Cumbitara, El Rosario y Leiva	1.959	5,63%	16.046	55.979	4,34%	52%	48%	66	1236	Agropecuaria, explotación de ganado bovino y especies menores, actividad minera.	Caña de azúcar, café, plátano, maíz, maní y frutales
<b>Centro</b>	Pasto, Nariño, La Florida, Yacuanquer, Tangua y Chachagüí	1.878	5,40%	360.966	105.365	28,09%	48%	52%	2924	6300	Agropecuaria, explotación de ganado bovino, porcino, equino y especies menores, actividad artesanal y comercial.	Papa, maíz, frijol, hortalizas

<b>Juanambú</b>	La Unión, San Pedro de Cartago, San Lorenzo, Arboleda y Buesaco	1.219	3,51%	20.285	64.54 3	5,11%	52%	48%	44	510	Agropecuaria, explotación de ganado bovino y especies menores, actividad artesanal.	Café, plátano, maíz, yuca, fique, caña de azúcar y frutales
<b>Río Mayo</b>	El Tablón de Gómez, Albán, San Bernardo, Belén, Colon-Génova, La Cruz y San Pablo	864	2,48%	26.921	77.34 1	6,28%	52%	48%	2376	1084	Agropecuaria, explotación de ganado bovino, porcino y especies menores, actividad artesanal e industrial.	Café, plátano, maíz, yuca, fique, caña de azúcar y frutales
<b>Guambuyaco</b>	Los Andes Sotomayor, La Llanada, El Tambo, El Peñol.	1.764	5,07%	15.038	28.82 1	2,64%	51%	49%	132	160	Agropecuaria, explotación de ganado bovino, porcino y especies menores, actividad minera.	Café, plátano, maíz, yuca, fique, caña panelera y frutales

Fuente. Este estudio



**Figura 1. Distribución por sub regiones departamento de Nariño**

Fuente: Gobernación de Nariño, 2012.

Muchas instituciones gubernamentales y académicas están hoy en día muy interesadas en el consumo energético de una región y en las posibles proyecciones de la demanda del recurso energético, entre ellas la UPME (Unidad de Planeación Minero Energética), IPSE (Instituto de planificación y promoción de soluciones energéticas para las zonas No interconectadas) a nivel nacional. A nivel regional también existe una preocupación por adelantar esfuerzos que permitan mejorar las condiciones de vida del departamento a través de una mejor prestación de los servicios



energéticos, instituciones como la universidad de Nariño con programas como El Plan de Energización Rural Sostenible (PERS Nariño) ha buscado la caracterización por uso y fuente de la energía en las zonas rurales del departamento.

En el año 2013 el PERS Nariño realizó un muestreo a las trece subregiones del departamento para investigar la fuente y los usos de la energía en los sectores residencial, comercial industrial e institucional, utilizando técnicas e instrumentos estadísticos para analizar la información.

El proyecto PERS Nariño ha utilizado un sistema de recolección de información primaria que permite realizar una clasificación de los consumos energéticos a niveles de residencia, comercial e industrial- institucional, además de que realiza una discriminación en trece subregiones, con características similares intra grupos y diferencias más grandes inter grupos.

El índice de ruralidad determina que el 74% de los municipios del departamento tienen uno que supera el 40 lo cual implica que gran parte de los asentamientos humanos están aislados geográficamente y por ende presentan dificultades a la hora de tener un desarrollo económico favorable, ya que en muchos de los casos ni siquiera cuentan con un sistema interconectado de suministro eléctrico.

Cabe mencionar que gran parte del departamento no posee infraestructura vial, marítima y/o fluvial adecuada, lo que dificulta enormemente el recorrido de esas distancias y que por lo tanto no se haya realizado una interconexión eléctrica, sino que la mayoría de sus habitantes utilicen otro tipo de alternativas para saciar sus necesidades de energía, que por lo general son de origen no renovable como derivados del petróleo y también de origen arbóreo.

### **3. Materiales y Métodos**

Esta investigación es de tipo descriptivo e inferencial bajo los enfoques cualitativo y cuantitativo utilizando el proceso de la metodología CRISP-DM para determinar las variables que más influyen o explican el consumo de energía eléctrica en el departamento de Nariño. Para el análisis cualitativo se categorizó la variable respuesta, consumo energía eléctrica de la vivienda, para analizar la asociación con las variables explicativas en los grupos: consumo eficiente (0,40kWh], consumo medio (40kWh, 90kWh] por debajo del consumo mínimo vital y consumo alto por encima de los 90kWh.

Esta metodología es pertinente ya que permite a través de una serie de pasos estructurados, comprender la situación problema o el entorno en donde ocurren los datos de la investigación, para empezar se hace un reconocimiento de los datos presentes y define si presentan problemas en cuanto a su estructura o recopilación, posteriormente se realiza una preparación de los datos lo cual incluye una limpieza o depuración de registros y atributos, para posteriormente manipularlos en la elaboración de un posible modelo, el cual se evalúa según los resultados que arroja en cuanto a la predicción en el comportamiento de las variables de estudio, si es necesario se repite el ciclo para depurar cada uno de los resultados que están implícitos en esta metodología, más adelante se describen en detalle la aplicación de esta metodología en la presente investigación.

#### **3.1. Comprensión del Problema**

##### **3.1.1. Contexto**

La Unidad de Planeación Minero Energética en el Plan de Expansión y Referencia 2015-2029, se pudo establecer que la incursión de fuentes renovables no convencionales de energía representa importantes beneficios para la demanda, dependiendo del porcentaje de penetración, brindando de

paso confiabilidad y desplazando generación más costosa. Particularmente, para el departamento de Nariño se debe tener en cuenta que las soluciones energéticas para las zonas no interconectadas pueden alternarse entre extensión de redes del SIN (Sistema de Interconexión Nacional) o soluciones aisladas, en cualquiera de los casos se debe realizar un estudio juicioso de las condiciones sociales, geográficas, ambientales y técnicas para avanzar en ampliación de cobertura en esta región. La presente investigación se basa en los estudios realizados por el PERS Nariño en el año 2013 acerca del consumo energético en las trece subregiones del departamento, para lo cual el PERS diseñó una serie de preguntas que se convirtieron en las variables de estudio del presente trabajo. La encuesta aporta información económica, social, geográfica, y sobre todo de consumo energético, dentro de este último ámbito se prioriza las fuentes de energía para usos domésticos (preparación de alimentos, iluminación y consumo de aparatos eléctricos), adicionalmente, se integran dos variables ambientales para el presente estudio, irradiación solar y velocidad del viento, para ser parte del análisis del consumo de energía.

Debido a que algunas regiones del departamento no cuentan con el servicio de interconexión eléctrica, es muy probable que las fuentes de energía provinieran en gran parte de energías no renovables o fósiles

### **3.1.2. Objetivo**

Partiendo del estudio previo, Plan de Energización Rural Sostenible para el Departamento de Nariño PERS, donde se aplicaron en 2013 un total de 2479 encuestas para viviendas del sector rural del departamento de Nariño, con un conjunto de 121 variables de tipo socioeconómico, de consumo, de condiciones de prestación del servicio, tenencia de equipos y disponibilidad de servicio de comunicación. Posterior a esto, se agregaron dos variables de tipo ambiental resultado del proyecto “Análisis de Oportunidades Energéticas con Fuentes Alternativas para el

Departamento de Nariño” ALTENAR. La presente investigación, propuso trabajar sobre este conjunto de información disponible para lograr el objetivo: “Caracterizar las variables socioeconómicas, demográficas y de uso de energía que explican mejor el consumo de energía en el sector rural del departamento de Nariño”.

### **3.2. Comprensión de los Datos**

Aquí se realiza la identificación de las variables presentes en la encuesta realizada por el PERS Nariño en el año 2013, la base de datos presenta inicialmente 2479 registros con un total de 121 variables codificadas con un número de pregunta tal como fue hecho en la encuesta, por ejemplo P.59 indica “Consumo aparatos eléctricos kilovatios hora al mes”. Los datos de esta investigación fueron encontrados por el PERS Nariño a través de encuestas realizadas en las trece sub regiones del departamento, para las cuales se utilizaron técnicas de muestreo que permitieron obtener datos representativos de la población, la cual no es muy homogénea teniendo en cuenta factores geográficos, económicos y sociales, además fueron obtenidos a través de un muestreo aleatorio, multietápico, estratificado, teniendo en cuenta las similitudes dentro de cada una de las subregiones que fueron definidas en el año 2012 a partir del plan de desarrollo, las cuales presentan rasgos distintivos entre sí.

#### **3.2.1 Metodología de recolección de información primaria**

Los datos obtenidos en esta encuesta presentan pruebas estadísticas que dan validez a la inferencia de los resultados a la población rural del departamento de Nariño.

Teniendo en cuenta el índice de ruralidad mayor a 40, se han excluido 8 cabeceras municipales, Pasto, Túquerres, Samaniego, Ipiales, Tumaco, Sandoná, La Unión y La Cruz.

El estudio fue realizado en los sectores residencial, institucional y comercial, pero para efectos de la actual investigación se enfatiza en el sector residencial en donde se tienen la mayor cantidad de observaciones.

Los datos provienen de encuestas realizadas en el área urbana y rural sin que esto implique una investigación de consumo energético urbano, pues los municipios seleccionados cumplen con el índice de ruralidad que se necesita para tal fin.

La mayoría de las encuestas fueron aplicadas en veredas (71,59%) y corregimiento (10,37%) y solo la décima parte en cabecera municipal (10,37%).

### 3.3 Descripción del Diccionario de Datos Inicial

Inicialmente la base de datos del PERS contiene información en 121 variables o campos, con 2479 registros, los cuales contienen información, socioeconómica, de ubicación, relacionadas con la vivienda, referente a telecomunicaciones, presencia de aparatos eléctricos específicos, condiciones del servicio de energía eléctrica, fuentes de obtención de energía y consumo de aparatos eléctricos en Kilovatios. Posteriormente se agregan las variables de ubicación como son latitud y longitud y variables ambientales como intensidad de viento e irradiancia global horizontal.

La base de datos original, tal como fue obtenida por el grupo PERS Nariño, se transformó remplazando los nombres de las variables que inicialmente se encontraban en forma de numero de pregunta y se cambiaron por su correspondiente nombre de acuerdo a su codificación tabla 2

**Tabla 2.**  
**Variables Encuesta PERS Nariño**

Denominación	Definición de las variables en la encuesta
<b>F. Expsubregión</b>	Factor de expansión por subregión
<b>F. Expmunicipio</b>	Factor de expansión por municipio
<b>Subregión</b>	Subregión
<b>Municipio</b>	Municipio
<b>Ubicación</b>	Datos de la ubicación de la vivienda

<b>Localidad</b>	Nombre de la localidad
<b>P.4</b>	Tipo de vivienda
<b>P4.1</b>	Cuál?
<b>P.5</b>	¿Cuántos hogares residen en la vivienda? (familiares o no)
<b>P.6</b>	Cuántas personas componen el hogar
<b>P.7.1</b>	Por favor indique cuántas salas tiene el hogar?
<b>P.7.2</b>	Por favor indique cuántas cocinas tiene el hogar?
<b>P.7.3</b>	Por favor indique cuántos comedores tiene el hogar?
<b>P.7.4</b>	Por favor indique cuántos baños tiene el hogar?
<b>P.7.5</b>	Por favor indique cuántas habitaciones tiene el hogar?
<b>P.7.6</b>	Por favor indique cuántos garajes tiene el hogar?
<b>P.7.7</b>	Por favor indique cuántas bodegas tiene el hogar?
<b>P.7.8</b>	Otro, Cuál?
<b>P.7.9</b>	Número
<b>P.8</b>	La vivienda es
<b>P.9</b>	El uso de la vivienda es exclusivamente
<b>P.9.4</b>	Otro, Cuál?
<b>P.10</b>	Cuál es el material predominante de las paredes exteriores?
<b>P.11</b>	Cuál es el material predominante de los pisos?
<b>P.12</b>	De dónde proviene generalmente el agua para consumo humano?
<b>P.13</b>	Cuánto tiempo o cuánta distancia recorre para llevar el agua a su casa?
<b>P.13.1</b>	Unidad de tiempo
<b>P.13.2</b>	Tiempo recorrido para llevar el agua a su casa (Horas)
<b>P.14</b>	La vivienda cuenta con alcantarillado
<b>P.15</b>	El servicio sanitario es
<b>P.16</b>	De qué forma eliminan las basuras?
<b>P.16.6</b>	Otra, cuál?
<b>P.17.1</b>	La vivienda Cuenta con teléfono fijo con línea
<b>P.17.2</b>	La vivienda cuenta con teléfono celular
<b>P.17.3</b>	La vivienda cuenta con equipo de radio para comunicaciones
<b>P.17.4</b>	La vivienda cuenta con internet
<b>P.18</b>	La vivienda utiliza servicio de energía eléctrica?
<b>P.19</b>	Cuenta con medidor de energía eléctrica?
<b>P.20</b>	Cuántos días a la semana tiene el servicio de energía eléctrica?
<b>P.21</b>	Cuántas horas al día tiene el servicio?
<b>P.22</b>	Podría facilitarnos un recibo de pago?
<b>P.23.1</b>	Periodo facturado
<b>P.23.2</b>	Valor pagado
<b>P.23.3</b>	Consumo en kWh
<b>P.23.4</b>	Estrato socioeconómico
<b>P.23.5</b>	Promedio de los últimos seis meses
<b>P.24.1</b>	¿Podría decirnos cuánto pagó por el servicio el último mes o un mes que recuerde? MES
<b>P.24.2</b>	Valor pagado

<b>P.25</b>	En una semana normal con qué frecuencia tiene interrupciones del servicio de energía eléctrica?
<b>P.26.1</b>	Si hay interrupciones indique Horas
<b>P.27</b>	De contar con energía eléctrica de manera confiable usted estaría dispuesto a pagar una tarifa mensual?
<b>P.28</b>	Por qué razón no estaría dispuesto a pagar?
<b>P.28.4</b>	Otra, Cuál?
<b>P.29</b>	Por la conexión o acometida al servicio de la red usted estaría dispuesto a pagar:
<b>P.30</b>	Cuanto estaría dispuesto a pagar mensualmente, una vez que cuente con el servicio?
<b>P.31</b>	Cuál es la fuente de iluminación principal en la vivienda?
<b>P.32.1</b>	¿En promedio cuántas horas al día utiliza esta fuente de iluminación? Horas
<b>P.32.2</b>	Minutos
<b>P.33.1</b>	Consumo lámparas incandescentes kWh mes
<b>P.33.1.1</b>	Consumo lámparas incandescentes Mcal Mes
<b>P.33.2</b>	Consumo lámparas ahorradores kWh mes
<b>P.33.2.1</b>	Consumo lámparas ahorradores Mcal Mes
<b>P.33.3</b>	Consumo lámparas fluorescentes kWh mes
<b>P.33.3.1</b>	Consumo lámparas fluorescentes Mcal Mes
<b>P.34</b>	Usa nevera o refrigerador?
<b>P.35</b>	Consumo refrigeración kWh mes
<b>P.35.1</b>	Consumo refrigeración Mcal Mes
<b>P.36</b>	Usa aire acondicionado o ventilador?
<b>P.37</b>	Consumo ambiente kWh Mes
<b>P.37.1</b>	Consumo ambiente Mcal mes
<b>P.38</b>	Usa algún sistema para agua caliente o calefacción
<b>P.39</b>	Consumo calefacción kWh mes
<b>P.39.1</b>	Consumo calefacción Mcal mes
<b>P.40</b>	En qué lugar de la vivienda cocinan
<b>P.41</b>	En general qué combustible usa principalmente para cocinar?
<b>P.41.8</b>	Otro, cuál?
<b>P.42</b>	Consumo estufas eléctricas kWhMes
<b>P.42.1</b>	Consumo estufas eléctricas McalMes
<b>P.44</b>	Utiliza una segunda estufa para cocinar?
<b>P.45</b>	Cada cuánto usa esta segunda estufa?
<b>P.45.2</b>	Unidad de medida
<b>P.45.3</b>	Frecuencia de uso al mes de la segunda estufa eléctrica
<b>P.46</b>	Podría indicar que combustible utiliza para la segunda estufa?
<b>P.46.8</b>	Otro, cuál?
<b>P.47</b>	Consumo de gas galones Mes
<b>P.47.1</b>	Consumo gas kWh mes
<b>P.47.1.1</b>	Consumo gas Mcal mes
<b>P.47.2</b>	Consumo carbón Kg/Mes
<b>P.47.2.1.1</b>	Consumo carbón Mcal mes

<b>P.48</b>	Podría indicar la cantidad de leña que utiliza un día para cocinar (Peso en Kgs)
<b>P.48.2</b>	A qué especie de madera pertenece la leña?
<b>P.49</b>	De dónde extrae la leña para cocinar?
<b>P.50</b>	¿Cada cuánto recolecta leña para cocinar? Cantidad
<b>P.50.1</b>	Frecuencia
<b>P.50.2</b>	Frecuencia de recolección al mes
<b>P.49.5</b>	Otro, cuál?
<b>P.51.1</b>	¿Cuánto tiempo o cuánta distancia recorre para llevar su leña a su casa? Tiempo
<b>P.51.2</b>	Unidad de tiempo
<b>P.51.3</b>	Tiempo que recorre para llevar la leña a la casa (Horas)
<b>P.52</b>	Dónde compra regularmente la leña?
<b>P.52.5</b>	Otra, cuál?
<b>P.53</b>	Qué tipo de estufa de leña tiene?
<b>P.53.5</b>	Otra, cuál?
<b>P.54.1</b>	¿A qué hora prende la estufa? 24 horas
<b>P.54.3</b>	La mantiene prendida todo el día?
<b>P.47.3</b>	Consumo de leña Kg Mes
<b>P.47.3.1</b>	Consumo de leña kWh Mes
<b>P.47.3.2</b>	Consumo de leña Mcal Mes
<b>P.55</b>	Utiliza el horno para preparar alimentos?
<b>P.56</b>	Qué combustible utiliza para el horno?
<b>P.56.7</b>	Otro, cuál?
<b>P.57</b>	Consumo hornos kWh mes
<b>P.57.1</b>	Consumo hornos Mcal mes
<b>P.59</b>	Consumo aparatos eléctricos kWh mes
<b>P.59.1</b>	Consumo aparatos eléctricos Mcal mes
<b>P.62</b>	Ingreso mensuales

Fuente. Este estudio

### 3.3.1 Limpieza de datos

Según (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004), eliminar toda la columna, es decir quitar el atributo para todos los ejemplos, es una solución ya que la proporción de datos nulos es tan alta que la columna no tiene arreglo.

Debido a que muchas de las 121 variables no aportan demasiada información, por tener la gran mayoría de sus registros vacíos, tal es el caso de las variables que indican valores adicionales a los valores más comunes, por ejemplo, cuando se pregunta que otro tipo de cuarto posee la vivienda aparte de habitaciones, cocinas, salas, o bodegas, comedores o baños, la variable “otro” cuarto



presenta 2402 registros vacíos es decir el 96,85%, por lo cual la variable se descarta para el presente estudio, así como otro tipo de variables las cuales no representan información significativa, con este mismo criterio de selección de variables solo se trabaja con las que aparecen en la siguiente tabla. En uno de los anexos se presentan las variables del estudio que no se tuvieron en cuenta.

Las variables consumo de energía en Kwh mes y Mcal mes que aparecen en muchas de las variables de la base inicial están correlacionadas por el factor de conversión, y consideración que es información redundante (1Mcal=1.116Kwh).

Las 55 variables trabajadas en esta investigación son las que aparecen en la tabla 3:

**Tabla 3.**  
**Descripción de variables en la investigación**

VARIABLE	TIPO	DESCRIPCIÓN
SUBREGIÓN	Cadena	Indica la subregión a la cual pertenece
UBICACIÓN	Cadena	Discrimina entre: vereda caserío, centro poblado, inspección de policía, corregimiento
TIPO_VIVIENDA	Cadena	Discrimina entre: casa, apartamento, otro
NUMERO_HOGARES_VIVIENDA	Numérico	Indica el número de familias por vivienda
NUMERO_PERSONAS	Numérico	Número de personas que componen el hogar
NUMERO_SALAS	Cadena	Número de salas que componen el hogar
NUMERO_COCINAS	Numérico	Número de cocinas por vivienda
NUMERO_COMEDORES	Cadena	Número de comedores por vivienda
NUMERO_BAÑOS	Numérico	Número de baños por vivienda
NUMERO_HABITACIONES	Numérico	Número de habitaciones por vivienda
GARAJES	Cadena	Número de garajes por vivienda
BODEGAS	Numérico	Número de bodegas por vivienda
ESTADO_VIVIENDA	Cadena	Indica si la vivienda es propia, prestada , en anticresis o arriendo
USO	Cadena	Indica la destinación de la vivienda
PARED	Cadena	Indica el material de las paredes de la vivienda
PISO	Cadena	Indica el material del piso de la vivienda

PROCEDENCIA_AGUA	Cadena	Indica la procedencia del agua que se consume
ALCANTARILADO	Cadena	Indica si presenta el servicio de alcantarillado
SANITARIO	Cadena	Indica si presenta el servicio de sanitario
BASURAS	Cadena	Indica el destino que se da a las basuras
CELULAR	Cadena	Indica si presenta el servicio celular
ENERGIA_ELECTRICA	Cadena	Indica si presenta el servicio de alcantarillado
MEDIDOR	Cadena	Presenta medidor de energía eléctrica
DIAS_ENERGIA	Numérico	Indica cuantos días a la semana cuenta con el servicio
HORAS_ENERGIA	Numérico	Indica cuantas horas al día cuenta con el servicio
RECIBO	Cadena	Indica su recibo de energía o no
CONSUMOKwh_AGRUPADO	Cadena	Indica el consumo eléctrico agrupado en Kilovatios hora
ESTRATO	Cadena	Estrato socioeconómico
PROMEDIO_CONSUMO	Cadena	Indica el promedio de consumo eléctrico
INTERRUPCIONES/SEMANA	Numérico	Indica la cantidad de interrupciones por semana
HORAS_INTERRUPCION	Numérico	Duración de la interrupción en horas
DISPUESTO_PAGAR	Cadena	Está o no está dispuesto a pagar por el servicio
ACOMETIDA_DISPUESTO_PAGAR	Cadena	Dinero dispuesto a pagar por la acometida eléctrica
DISPUESTO_APAGAR_HAYA_SERVICIO_AGRUPADO	Cadena	Dinero dispuesto a pagar por la interconexión
FUENTE_ILUMINACION	Cadena	Tipo de iluminación que presenta en el hogar
LAMPARAS_Kwh_MES_AGRUPADO	Cadena	Consumo de lámparas en Kilovatios al mes
REFRIGERADOR	Cadena	Tiene refrigerador
REFRIGERADORKwh_MES	Cadena	Consumo de refrigerador en Kilovatios
AIRE_ACONDICIONADO	Cadena	Tiene aire acondicionado
AIRE_ACONDICIONADO_Kwh_MES	Cadena	Consumo aire acondicionado en kilovatios al mes
CALEFACCION	Cadena	Tiene Calefacción
CALEFACCION_Kwh_MES	Cadena	Consumo calefacción en kilovatios al mes

LUGAR_COCINA	Cadena	Especifica si el lugar de la cocina es exclusivo solo para cocinar
COMBUSTIBLE_COCINAR	Cadena	Indica el tipo de combustible usado para cocinar
ESTUFA_Kwh	Cadena	Consumo de la estufa en kilovatios al mes
GAS_GALONES_MES	Cadena	Cantidad de gas en galones utilizados al mes
GAS_KwH_MES	Cadena	Consumo de gas en kilovatios al mes
LEÑA_Kg/DIA	Cadena	Consumo de leña en kilogramos al día
LEÑA_Kg_MES	Cadena	Consumo de leña en kilogramos al mes
LEÑA_KW_MES	Cadena	Consumo de leña en kilovatios al mes
APARATOS_ELECTRICOS_KwH_MES	Cadena	Consumo de aparatos eléctricos en kilovatios al mes
VIENTO	Cadena	Intensidad del viento
INGRESOS	Cadena	Ingresos económicos de las familias encuestadas
SOL	Cadena	Intensidad de sol
VALOR PAGADO O RECORDADO AGRUPADO	Cadena	Valor pagado por el consumo de energía eléctrica

Fuente. Este estudio

En seguida se comparan las diferentes columnas de la base de datos con el documento en donde se describen las preguntas y el formato de la encuesta, identificando grupos de variables de tipo socioeconómico, relacionadas con la vivienda, de consumo energía eléctrica y presencia de algún electrodoméstico en particular.

### 3.3.2 Transformación.

Algunas de las variables de la presente investigación son de tipo continuo, sobre todo las que describen el consumo de aparatos eléctricos, como son “consumo de aparatos eléctricos en kilovatios al mes” y “consumo de aparatos eléctricos en Megacalorías al mes”, cabe aclarar que estas variables están asociadas por su factor de conversión.

Inicialmente con las variables de tipo continuo se realizó un análisis de correlación, y posteriormente, se procedió a categorizarlas, para poder analizar la asociación entre variables.

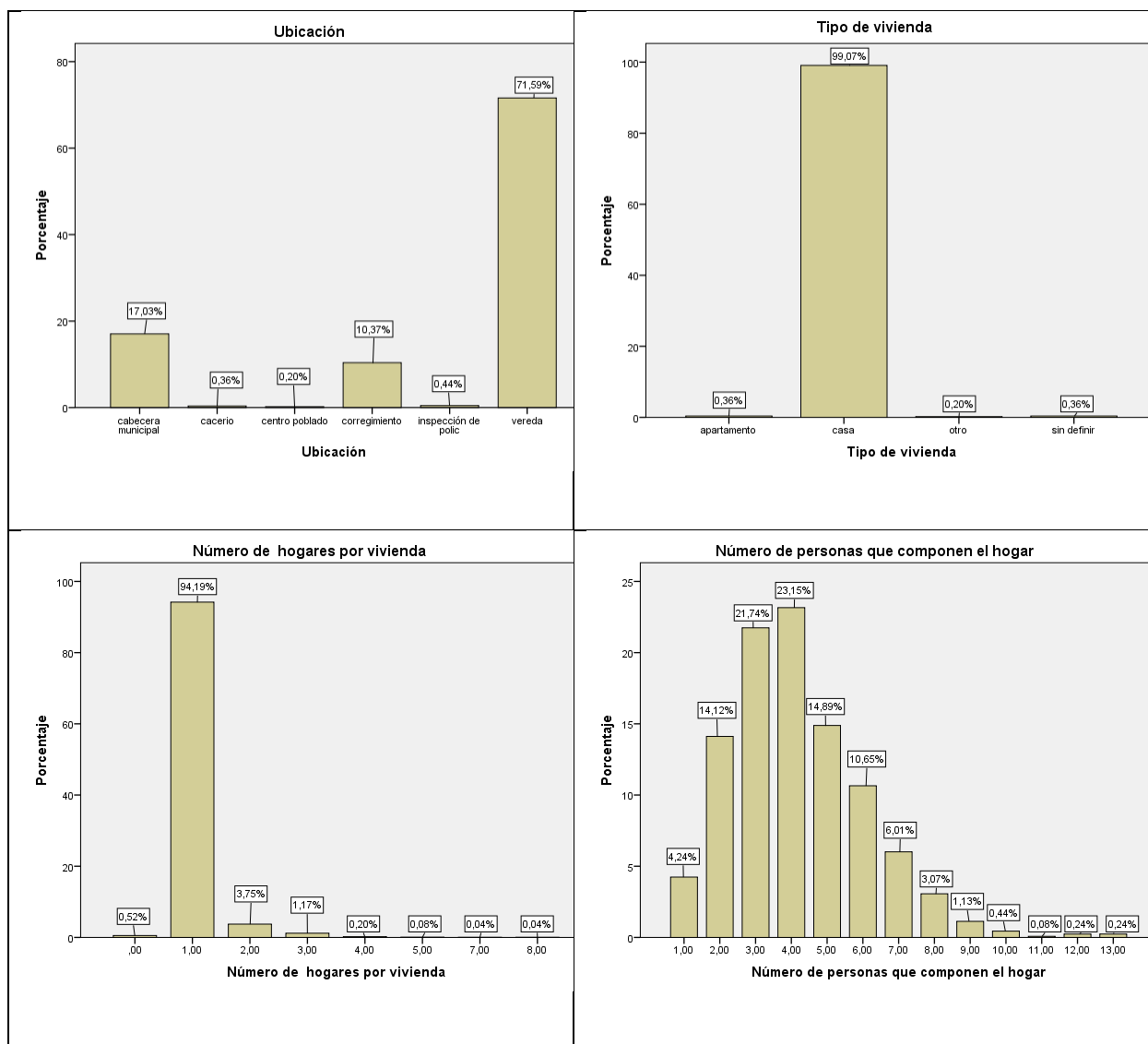
También se realizó la suma de variables que representan el consumo de energía eléctrica, entre ellas consumo en: refrigeración, aire acondicionado, calefacción, consumo tanto de lámparas incandescentes como ahorradoras, y también la variable consumo de aparatos eléctricos en kilovatios hora, para resumirse en la variable “consumo eléctrico total”, la cual posteriormente fue categorizada, cabe aclarar que las anteriores variables son de tipo continuo y todas están en la misma unidad de medida. La variable que resulta de la sumatoria se conoce como consumo eléctrico total pues resume el consumo de energía de todos los electrodomésticos del hogar como son televisores, equipos de sonido, planchas, licuadoras etcétera, y los aparatos que más consumo representan, como son neveras, aire acondicionado, aparatos de calefacción y lámparas eléctricas de todo tipo, es esta la variable respuesta del estudio, que para el caso del modelamiento fue categorizada, teniendo en cuenta el consumo eficiente de 40 kilovatios hora mensual aproximadamente.

### **3.4 Resultados**

Aquí se tratan los resultados obtenidos en el análisis descriptivo univariado, el análisis bivariado y por último el análisis multivariado y el modelo predictivo de consumo eléctrico

#### **3.4.1 Análisis univariado.**

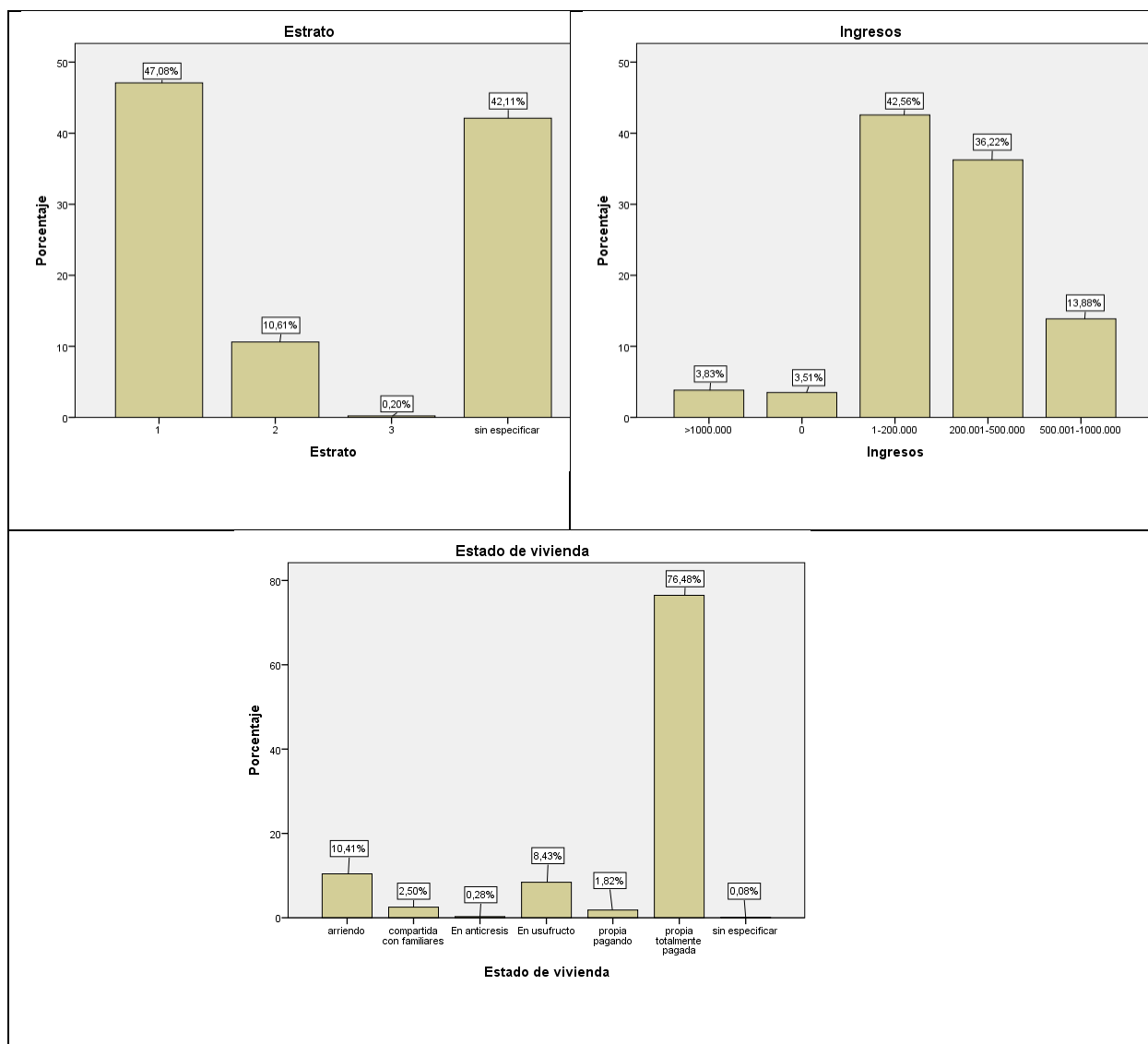
En esta sesión se realiza una descripción de las variables más relevantes de la investigación. Las siguientes graficas se realizaron teniendo en cuenta agrupaciones de variables como son las relacionadas con la vivienda, las variables socioeconómicas, las de consumo eléctrico, las de condiciones del servicio eléctrico y las de consumo de energía de aparatos eléctricos.



**Figura 2. Características consumo- Composición hogares**

Fuente. Este estudio

Las gráficas de la figura 2 indican el porcentaje de encuestas realizadas en las diferentes subregiones, y se observa que la gran mayoría se realizaron en veredas con un 71,59% del total, seguido de los centros poblados con un 17,03% y los corregimientos con un 10,37%, los demás lugares encuestados apenas suman el 1% del total de las encuestas. La gráfica de la composición de los hogares encuestados, indica que el 63,25 % de los hogares está conformado por máximo cuatro personas y el 36,75 % por 5 hasta 13 personas por hogar.

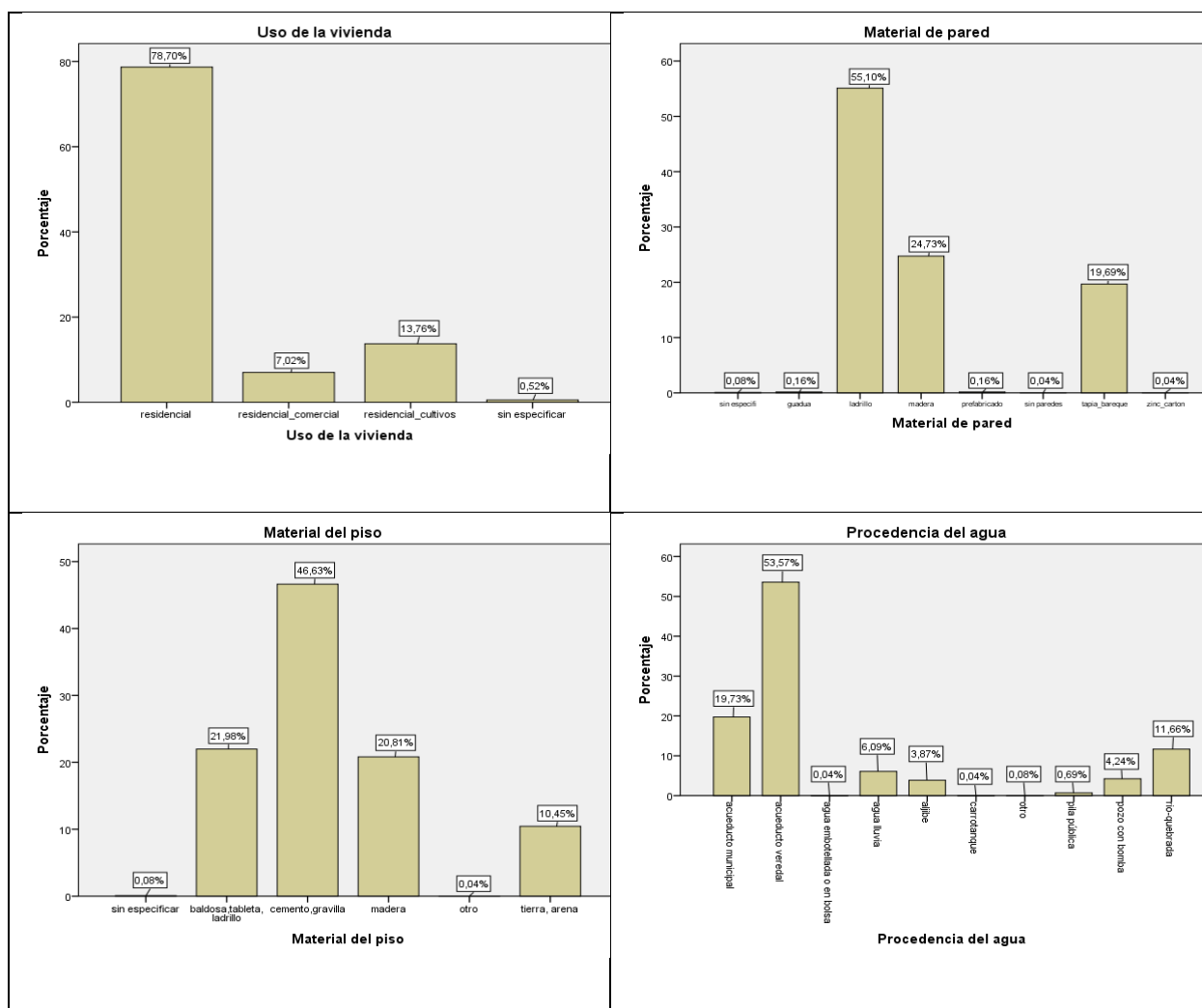


**Figura 3. Consumo hogares-Estrato, ingresos, estado vivienda**

Fuente. Este estudio

Dentro de las regiones consideradas rurales, el 47,08% pertenecen al estrato 1, el 10,61% al estrato 2, apenas el 0,20% al estrato 3 y un gran porcentaje de 42,11% no especifica el estrato al cual pertenece. En cuanto a la condición de la vivienda, el 76,46% tiene vivienda propia totalmente pagada, el 10,41% vive en arriendo, el 8,43% vive en usufructo, el 2,50% comparte la vivienda con familiares, respecto a los ingresos por hogar, el 42,56% posee un ingreso que va máximo hasta

los \$ 200.000, el 38,22% ingresos mensuales entre \$ 200.001 y \$ 500.000, el 13,88% entre \$500.001 y \$1.000.000 y solo el 3,83% ingresos superiores a \$1.000.000



**Figura 4. Uso vivienda, material de pared, material piso, procedencia de agua**

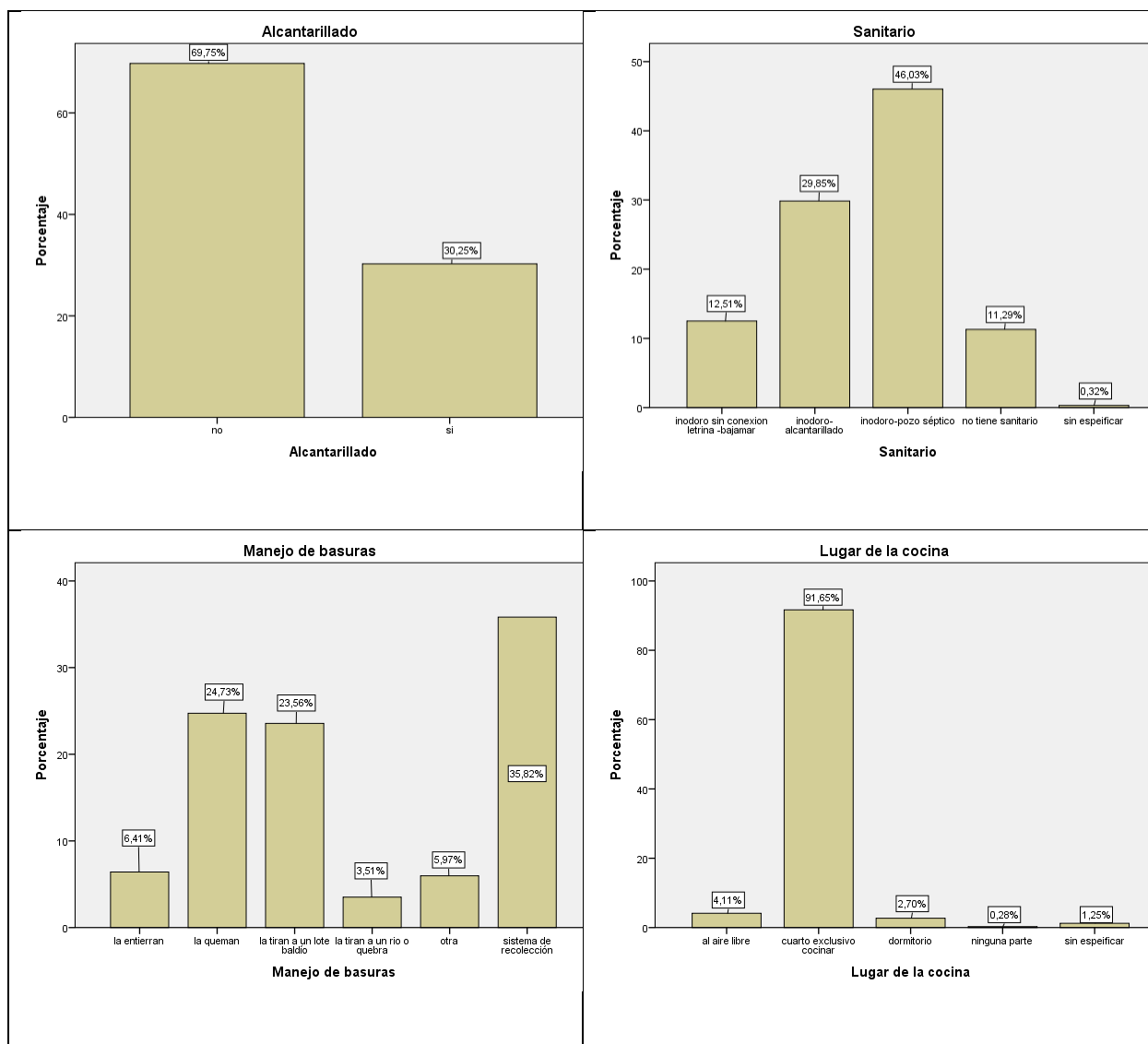
Fuente. Este estudio

El 78,70 % de las familias utiliza su vivienda para uso exclusivo residencial, mientras que el 13,76% de los encuestados afirma que aparte del uso residencial también la utilizan como bodega o de alguna manera guardar los insumos para los cultivos y un 7,02% utiliza la vivienda con un uso residencial-comercial; respecto al material de las paredes de la vivienda el 55,10% las tiene fabricadas con ladrillo, el 24,73% las tiene de madera, el 19,69% tiene la pared de tapia-banque y el 0,44% en otros materiales y el 0,4% vive sin paredes, respecto a el material de piso el 46,63%

de las viviendas tiene el piso de cemento-gravilla, el 21,98% de baldosa, tableta o ladrillo y el 20,81% de madera, el 10,45% de tierra o arena y el 0,8% no especifica.

En lo que respecta a la procedencia de agua, la mayoría el 53,57% de las familias poseen un acueducto veredal, seguido con un 19,73% que poseen un acueducto de destinación municipal, el 11,66% obtiene el agua de los ríos o quebradas, el 6,09% consume agua lluvia, el 8,11% recurre a los aljibes o pozos con bomba, el 0,69% utiliza una pila pública, el 0,04% utiliza agua de un carro tanque y el 0,08% de manera no especificada.





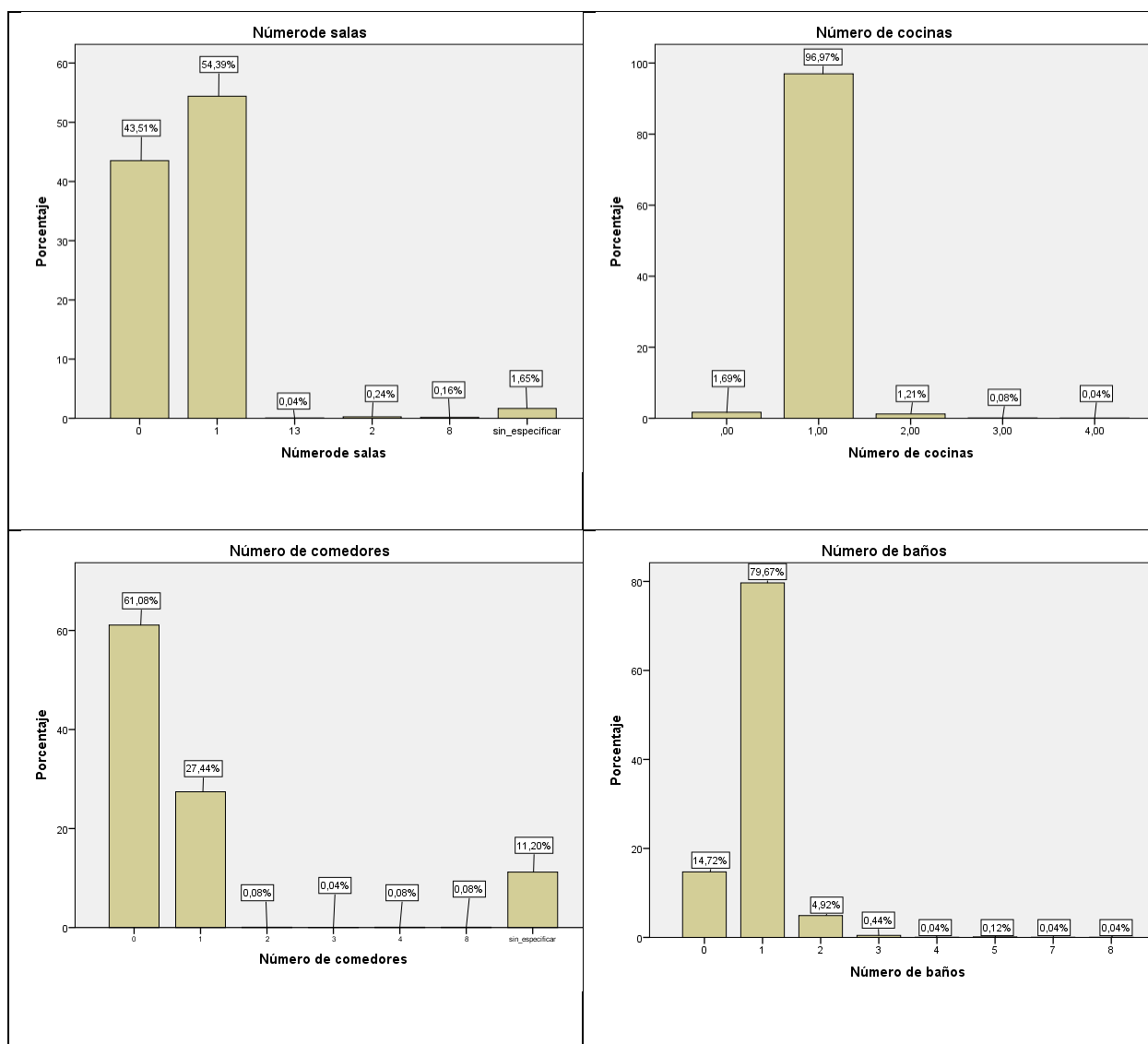
**Figura 5. Alcantarillado, sanitaria, manejo de basuras, lugar de cocina**

Fuente. Este estudio

El 69,75% de las familias encuestadas no cuenta con alcantarillado debido a que en la zona rural las casas se encuentran dispersas unas de otras, solo el 30,25% cuenta con este servicio, el cual aparece en las cabeceras municipales o centro poblados y en algunos caseríos; respecto al servicio sanitario, el 46,03% de las familias tiene inodoro conectado a pozo séptico, el 29,85% tiene inodoro conectado al alcantarillado, el 12,51% posee letrina o inodoro con conexión a baja mar, el 11,29% no tiene sanitario y el 0,32% no especifica.

Respecto a el manejo de las basuras, el 35, 82% de las familias cuentan con el sistema de recolección, el 24,73% la queman, el 23,56% tiran la basura a un lote baldío, el 6,41% la entierran, el 3,51% la tiran a un rio o quebrada y el 5,97% tiene otra disposición final.

En lo que refiere al lugar que se tiene destinado para la cocina, el 91,65% de los hogares posee un cuarto exclusivo para tal fin, el 4,11% cocina al aire libre, el 2,70% cocina en el dormitorio, el 1,53% no especifica o posee algún lugar determinado.



**Figura 6. Número de: salas, cocinas, comedores baños**

Fuente. Este estudio

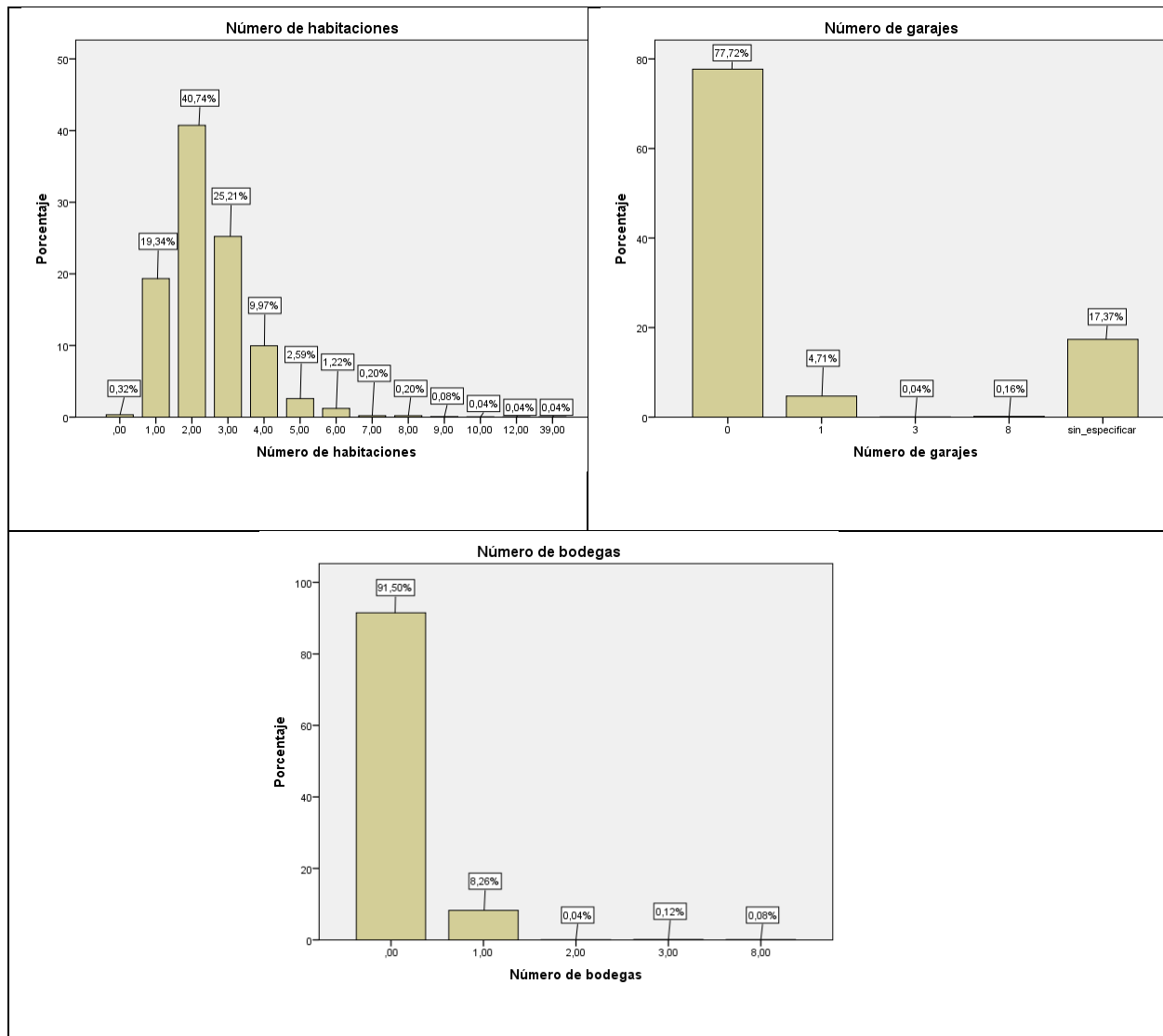
La cantidad de cuartos en las viviendas de los hogares encuestados se dispone de la siguiente manera:

El número de salas que presenta cada vivienda, se observa que un 43,51 % no posee sala en su casa y un poco más de la mitad (54,39%) presenta una sala.

En cuanto al número de cocinas el 96,97 % de los encuestados posee una, el 1,69 % no posee ninguna y el 1,33 % tiene dos cocinas o más.

El número de baños que existen en las viviendas presenta la siguiente distribución, en el 79,67% de los casos hay un baño, el 14,72% no tiene baño, el 4,92 % tiene 2 baños y el 0,68% tiene 3 baños o más.

En referencia a comedores por vivienda, el 61,06 % no lo tiene y el 27,44 % sí.



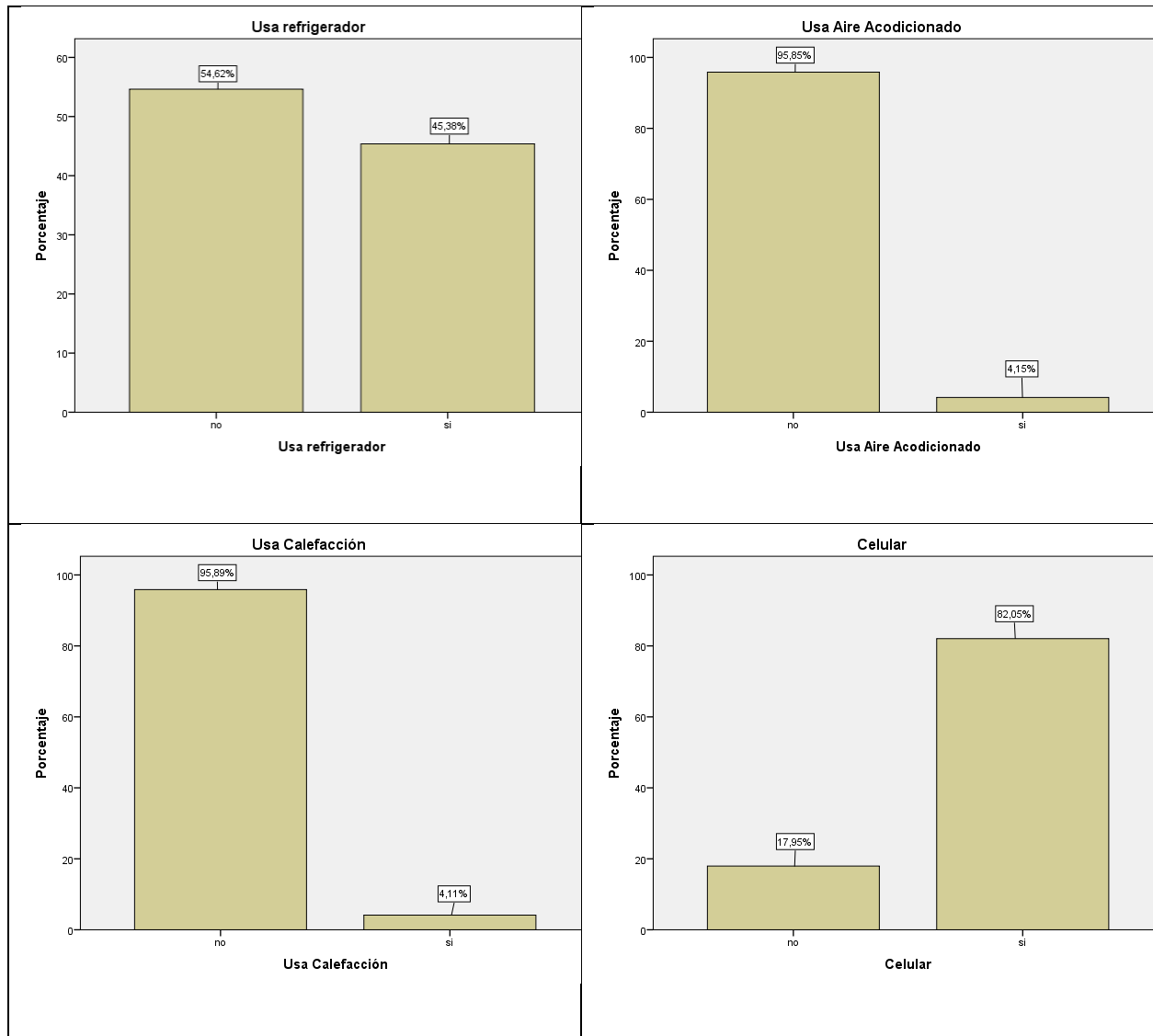
**Figura 7. Número de: habitaciones, garajes, bodegas**

Fuente. Este estudio

La grafica muestra el número de habitaciones por vivienda, se observa que el 60.4 % de total tienen dos o menos habitaciones y 39.6% más de 2 habitaciones

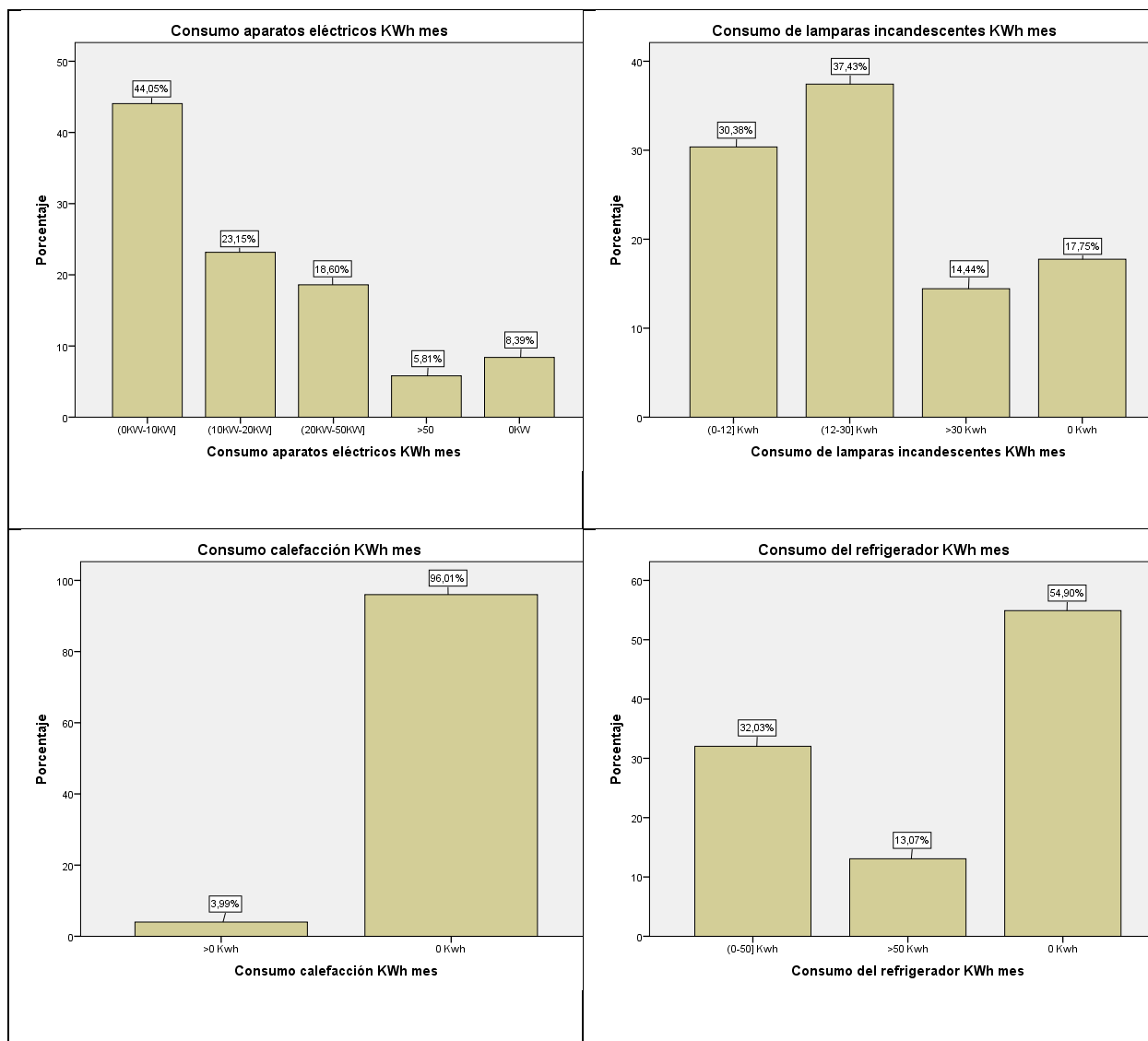
El 77,72% de los encuestados no tiene garaje en su vivienda, solo el 4,71% dispone de uno y el 0,20% dispone de más de un garaje, el 17,37% no lo especifica

El 91,50% no posee bodegas, el 8,26% posee una y el 0,24% posee más de una

***Aparatos eléctricos:*****Figura 8. Uso aparatos eléctricos**

Fuente. Este estudio

En cuanto al uso de aparatos de mayor consumo eléctrico a excepción del celular, se observa que el 54,62% hace uso de refrigerador frente al 45,38% que no lo hace; el aire acondicionado solo es usado por el 4,15%; calefacciones utilizada por el 4,11%; uno de los aparatos de telecomunicación más utilizados es el celular el cual es usado por 82,05% frente al 17,95% que no lo hace.

**Consumo energía eléctrica:****Figura 9. Consumo en viviendas**

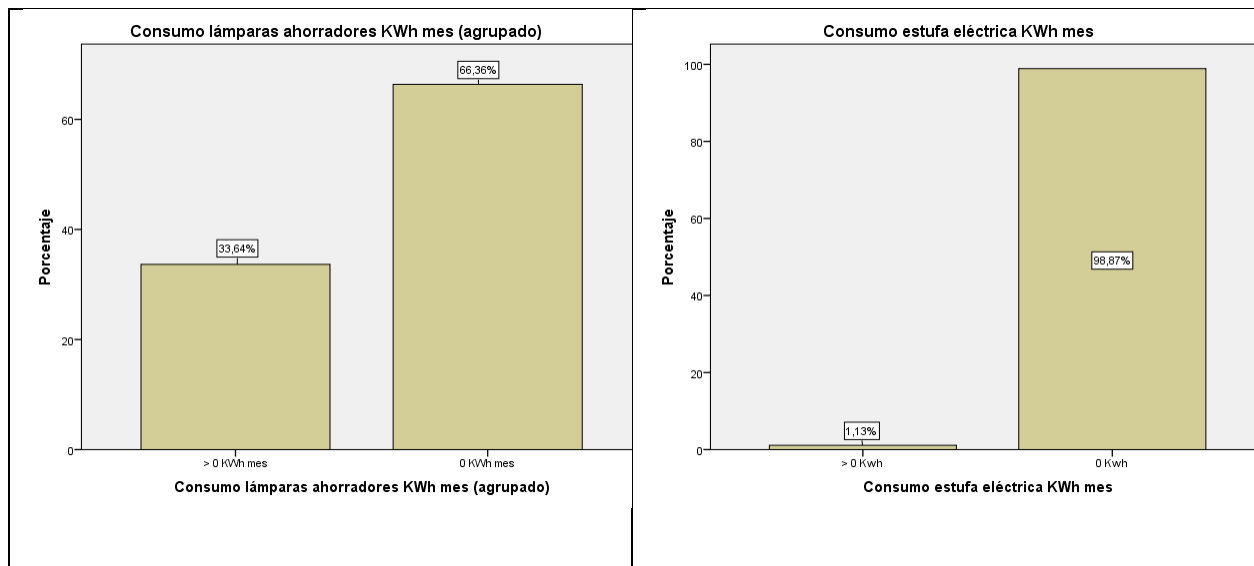
Fuente. Este estudio

En cuanto al consumo de aparatos eléctricos, el 8,39% de los hogares no registra consumo eléctrico, el 44,05% tiene un consumo entre 0 y 10 kilovatios hora al mes, el 23,15% consume entre 10 y 20 kWh al mes, el 18,60% entre 20 y 50 kWh al mes y el 5,81% consumo mayor a 50 kWh mes.

El consumo de lámparas incandescentes en kilovatios hora al mes está distribuido así: 0 kWh 17,75%; (0-12] kWh 30,38%; (12-30] kWh 37,43% y mayor a 30 kWh 14,44%.

El consumo de la calefacción tiene 96,01% cuyo consumo es 0 kWh y 3,91% un valor mayor.

El refrigerador tiene un consumo distribuido así: 54,90% no presenta consumo de este aparato, el 32,03% un consumo menor a 50 kilovatios hora al mes y el 13,07% un consumo mayor a 50 kWh mes.

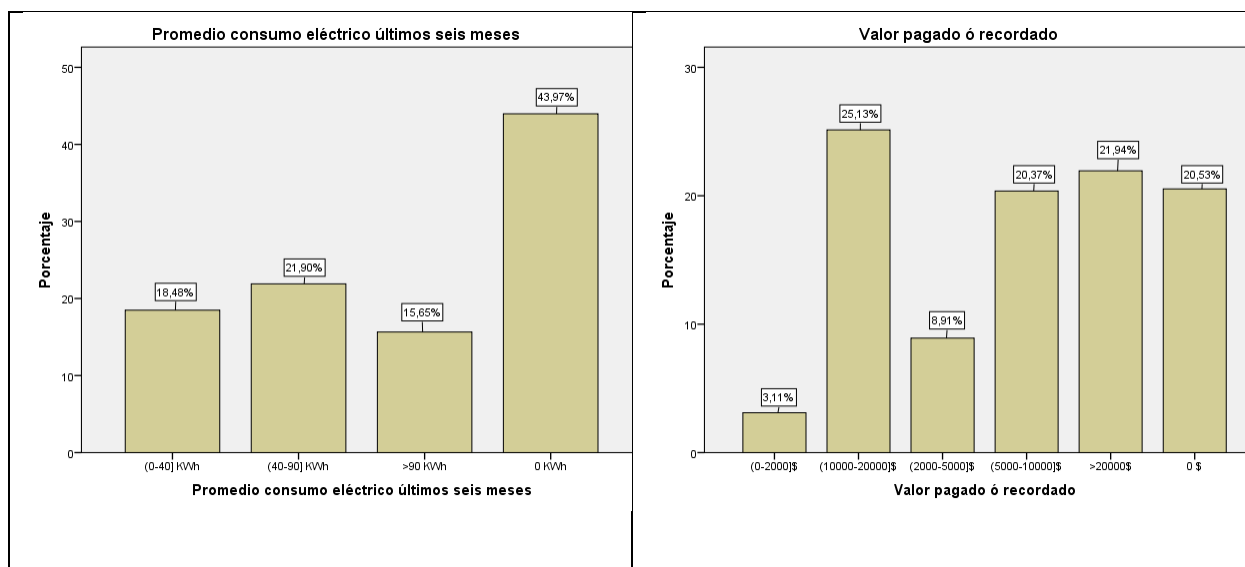


**Figura 10. Consumo en viviendas: Lámparas ahorradoras y estufa eléctrica**

Fuente. Este estudio

El consumo de lámparas ahorradoras es apenas de la tercera parte con un 33,64% que presentan consumo diferente de 0 kWh mes, frente a 66,36% que no hacen uso de este aparato ahorrador.

El 98,87% de los encuestados no utiliza la estufa eléctrica frente al 1,13% que registra un consumo mayor a 0 kW h mes.



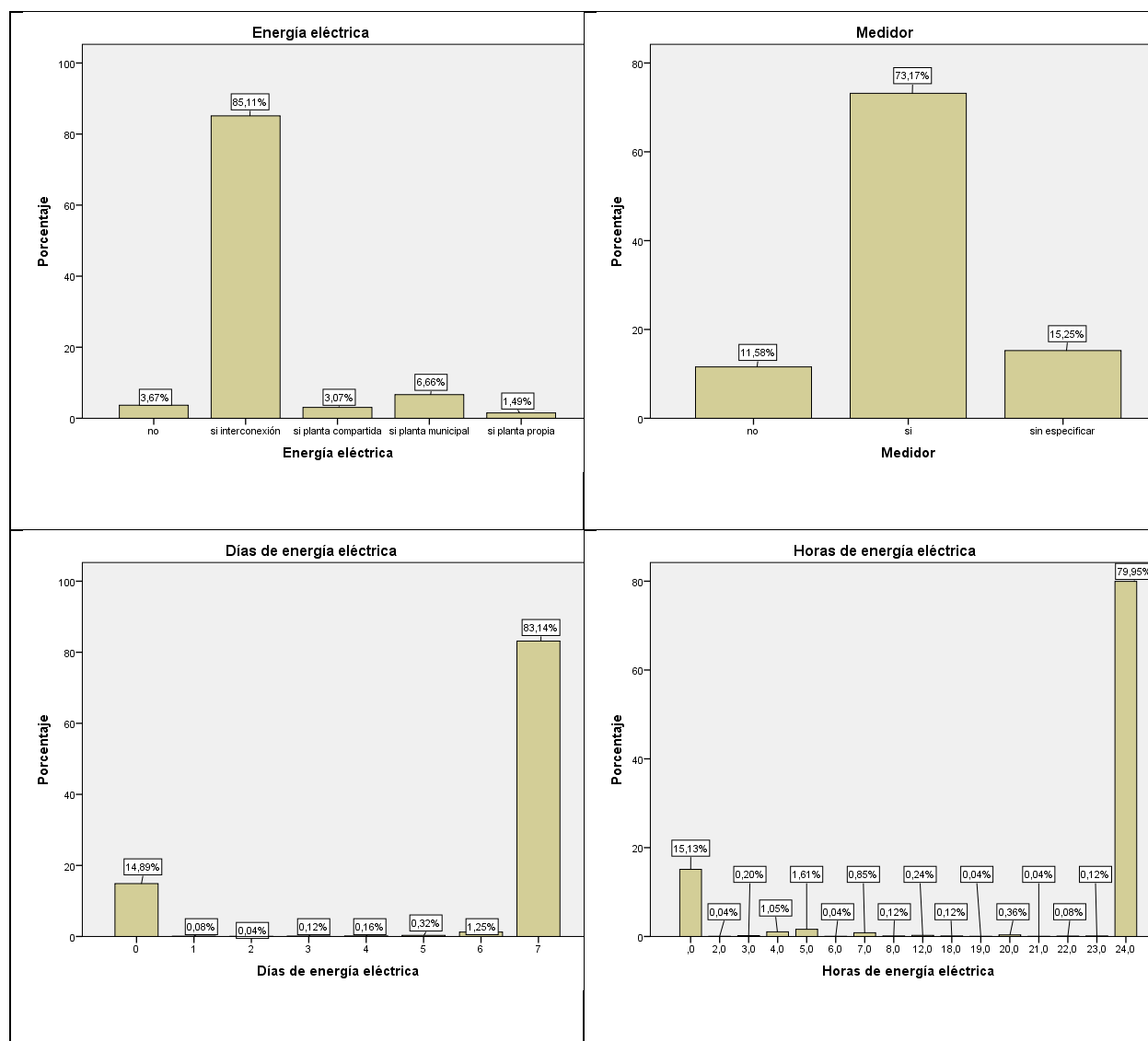
**Figura 11. Promedio consumo eléctrico y valor pagado o recordado**

Fuente. Este estudio

Un referente del consumo eléctrico de las familias es el promedio de los últimos 6 meses, el cual está distribuido así: el 43,97% no registra consumo, el 18,46% tiene un consumo entre 0 y 40 kilovatios hora al mes, el 21,90% entre 40 y 90 kWh mes y el 15,65% un consumo mayor a 90 kWh mes.

Por algún motivo, entre ellos tener planta eléctrica o simplemente no tener interconexión eléctrica, un gran porcentaje que no registran consumo tienen un valor pagado diferente de cero, el 20,53% de los encuestados no pagan dinero, el 3,11% paga entre \$0 y \$2000; el 8,91% paga entre \$2000 y \$5000; el 20,37% paga entre \$5000 y \$10000; el 25,13% paga entre \$10000 y \$20000 y el 21,94% paga más de \$20000.



*Condiciones del servicio de energía eléctrica:***Figura 12. Condiciones del servicio de energía eléctrica**

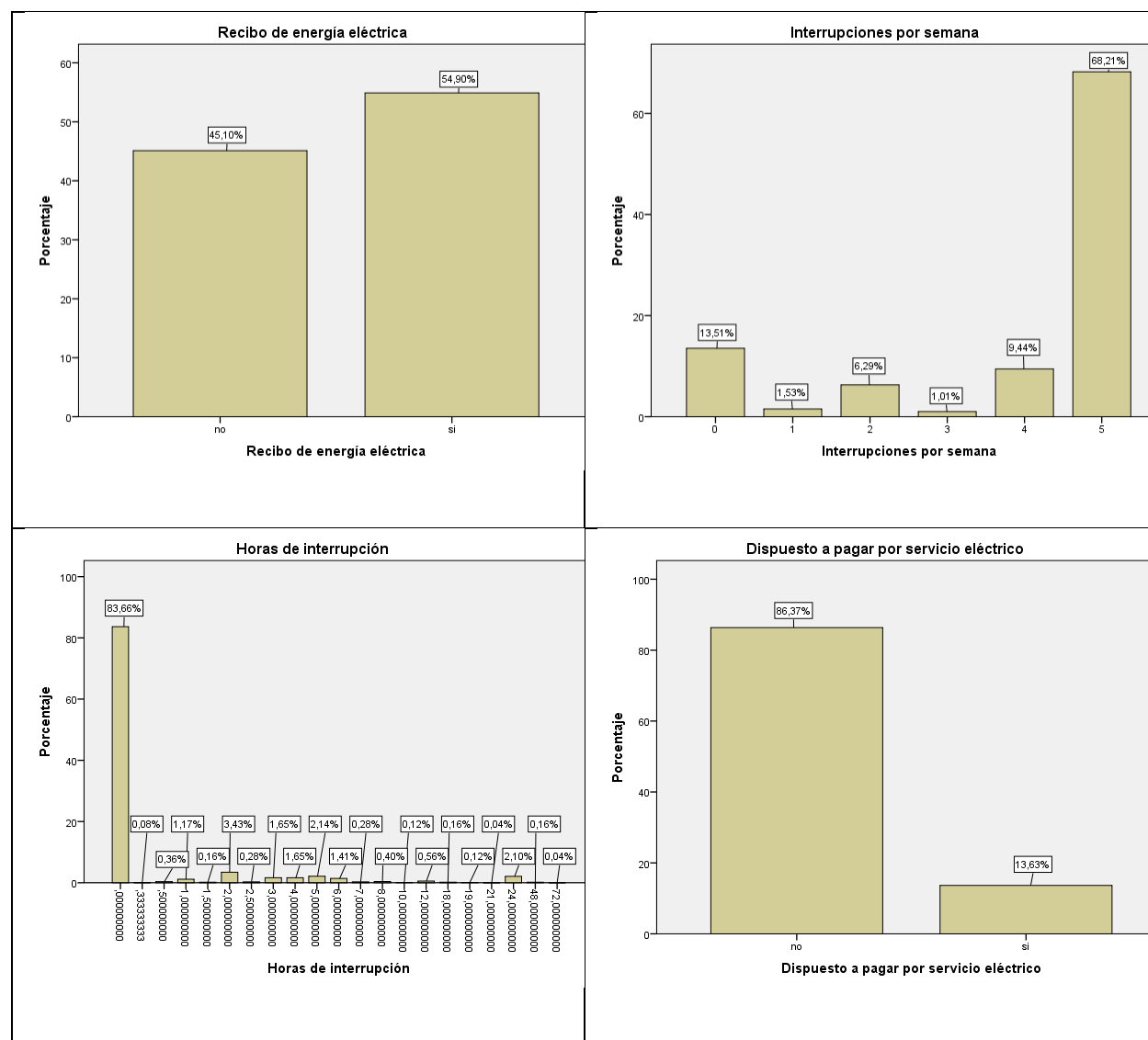
Fuente. Este estudio

En cuanto a las condiciones de la energía eléctrica, el 85,11% cuenta con el servicio de interconexión, el 6,66% con planta municipal, el 3,07% con planta compartida, el 1,49% con planta propia y el 3,67% no cuenta con el servicio.

En cuanto al medidor de energía no todos los que cuentan con interconexión presentan medidor, pues solo el 73,17% lo tiene es decir que el 11,94% de la muestra encuestada cuenta con el servicio

de interconexión, pero no tiene medidor instalado; el 11,58% no tiene medidor y el 15,25% no lo especifica.

Con respecto a los días de servicio eléctrico, el 83,14% cuenta con los 7 días de la semana, el 1,97% entre 1 y 6 días de servicio y el 14,89% ningún día de la semana; y en relación con la cantidad de horas de servicio, el 79,95% cuenta las 24 horas del día, el 15,13% o horas de servicio al día y el 4,92% entre 2 y 23 horas de servicio al día.

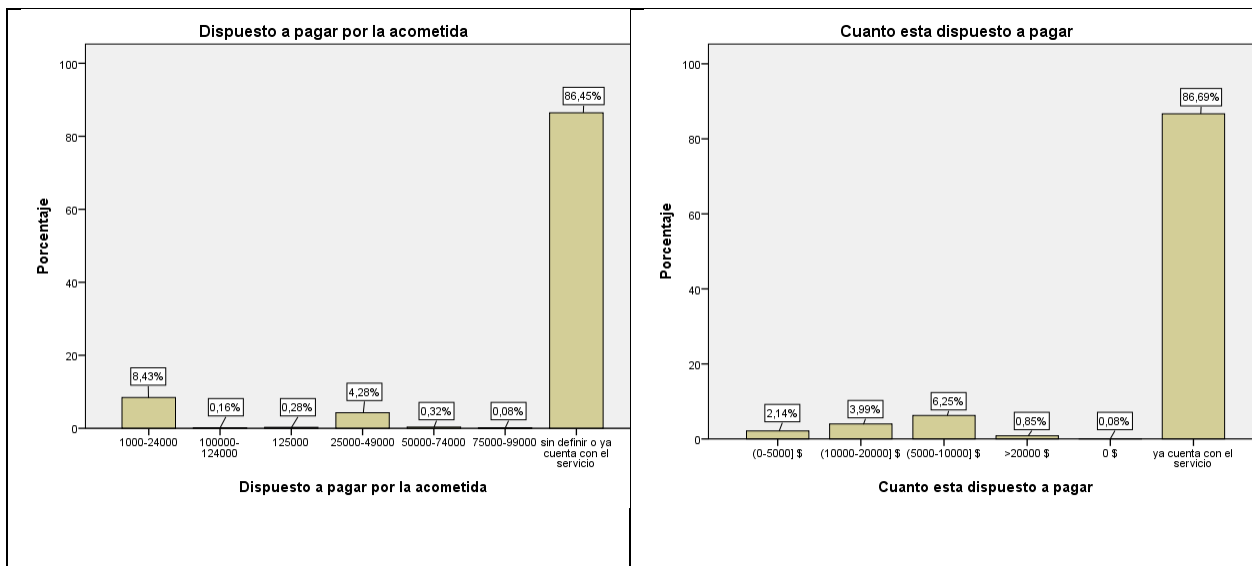


**Figura 13. Interrupciones del servicio, recibo energía, disposición a pagar por el servicio**

Fuente. Este estudio

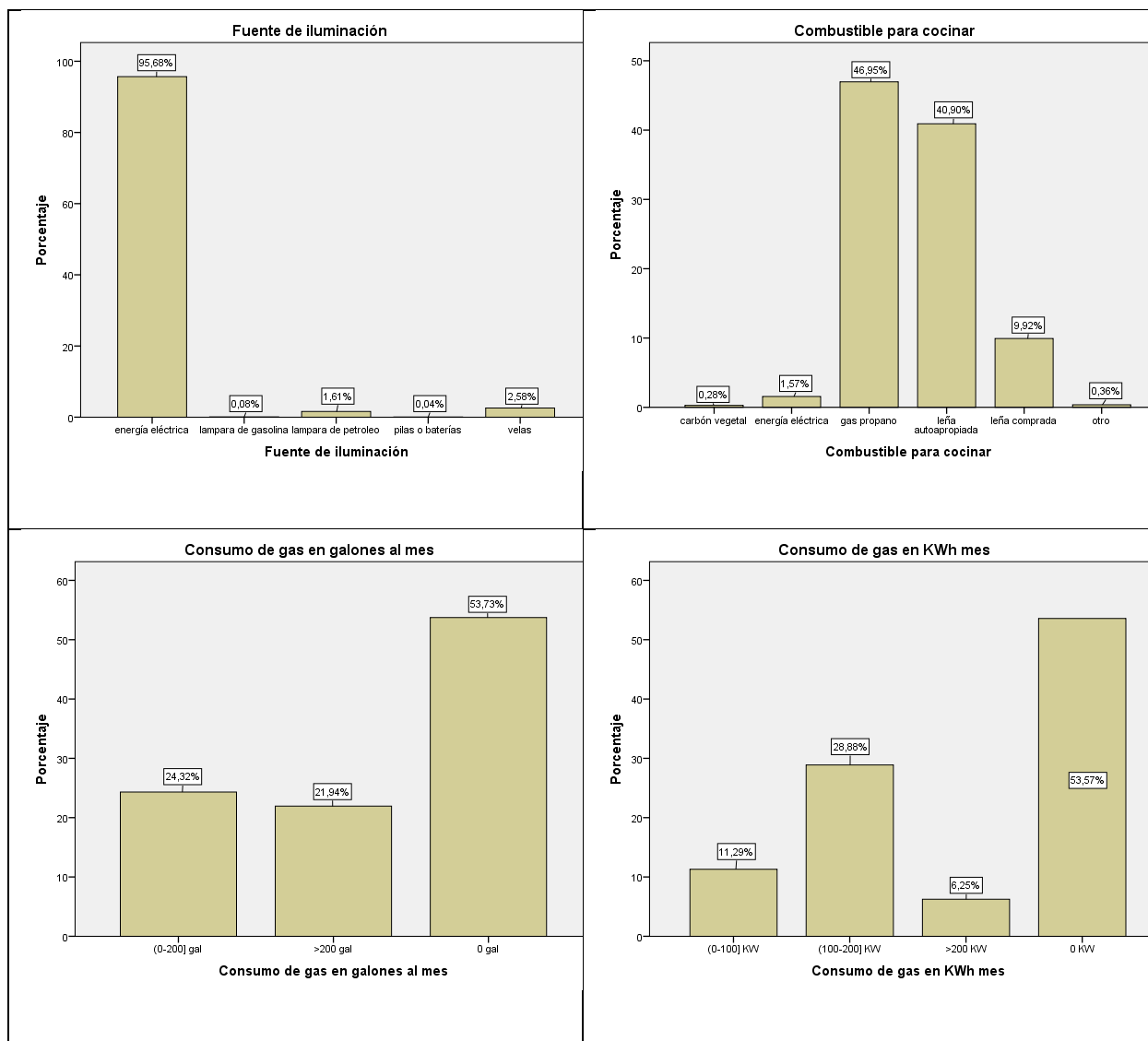
El porcentaje de personas que tienen recibo de energía eléctrica es: 45,10% no presenta recibo frente a un 54,90 5 que si lo tiene.

En cuanto al número de interrupciones o cortes de energía por semana, la encuesta está así: el 68,21% presenta 5 interrupciones, el 13,51% 0 interrupciones y 18,27% entre 1 y 4 interrupciones por semana. En horas de interrupción, el 83,66% no presenta ninguna hora y el 16,34% restante presenta una duración de media hora hasta 72 horas de interrupción.



**Figura 14. Dispuesto a pagar por acometida, Cuanto está dispuesto a pagar**

Fuente. Este estudio

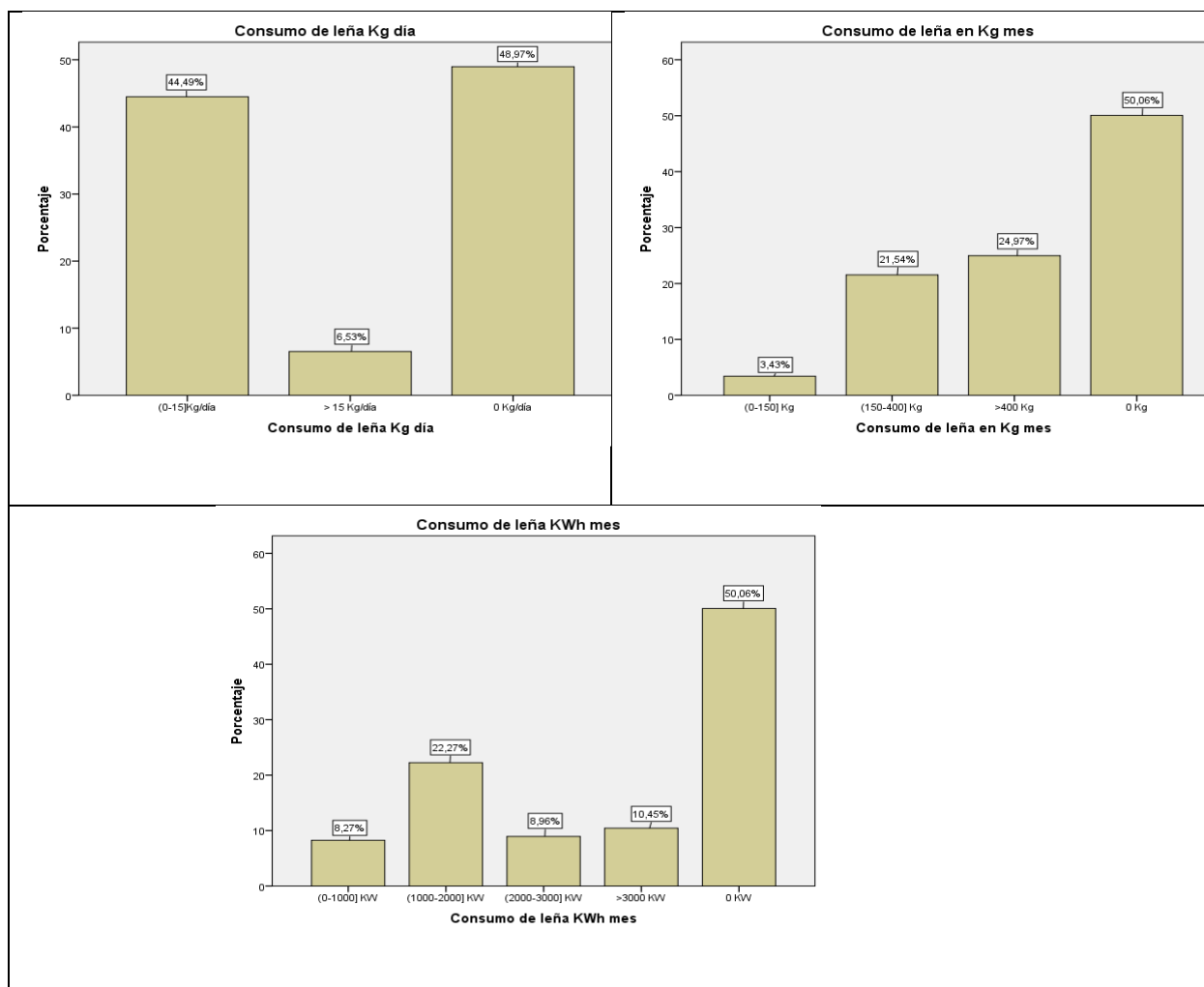
**Fuentes de energía:****Figura 15. Fuentes de energía**

Fuente. Este estudio

Las fuentes utilizadas para la iluminación de la vivienda están distribuidas así: el 95,65% utiliza energía eléctrica, el 2,58% utiliza velas, el 1,61% utiliza lámpara de petróleo, el 0,08% lámpara de gasolina y el 0,04% utiliza pilas o baterías. Respecto a la principal fuente de combustible utilizado para cocinar, el 46,95% utiliza gas propano, el 40,90% leña auto apropiada, el 9,92% leña comprada, el 1, 57% energía eléctrica, el 0,28% carbón vegetal y el 0,36% otra fuente diferente.

La utilización de gas por las familias encuestadas está distribuida así: el 53,73% no alcanza a utilizar ningún galón de gas al mes, el 24,32% utiliza entre 0 y 200 galones al mes, y el 21,94% utiliza más de 200 galones al mes.

Con respecto al consumo de gas en kWh al mes, el 53,73% no presenta consumo, el 11,29% entre 0 y 100kWh, el 28,88% entre 100 y 200 kWh, y el 6,25% más de 200kWh al mes.



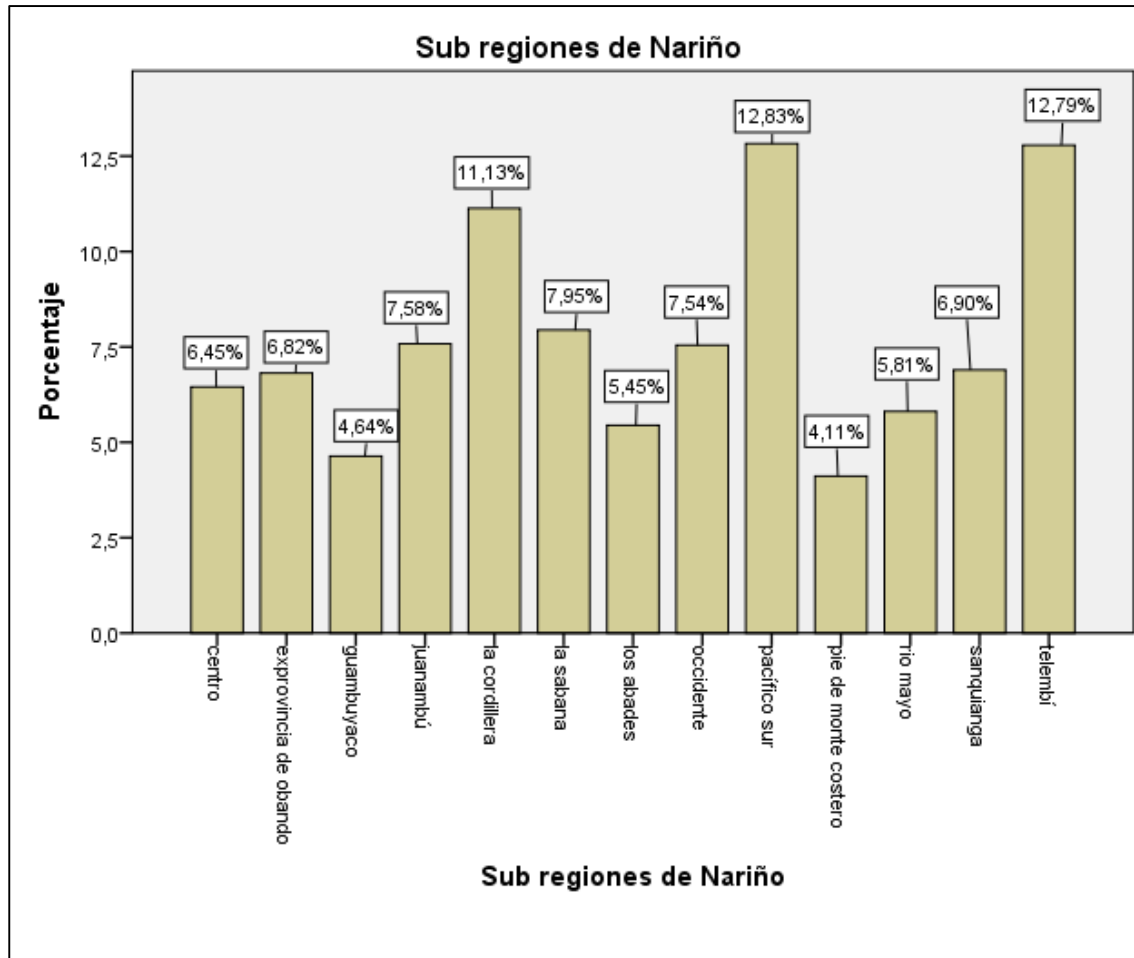
**Figura 16. Consumo para cocción**

Fuente. Este estudio

El consumo de leña en kilogramos al día se encuentra distribuido así: el 48,97% 0 Kg, el 44,49% entre 0 y 15Kg y el 6,53% más de 15 Kg.

El consumo de leña al mes está así: 50,06% no consumen ningún kilogramo al mes, el 3,43% entre 0 y 150 kg, el 21,54% entre 150 y 400 Kg, y el 24,97% más de 400 Kg al mes.

Con respecto al consumo de leña en kilovatios hora al mes, se tiene que el 50,06% no consume ningún kWh, el 8,27% entre 0 y 1000 kWh, el 22,27% entre 1000 y 2000 kWh, el 8,96% entre 2000 y 3000kWh y el 10,45% más de 3000kWh al mes.



**Figura 17. Distribución de encuestas**

Fuente. Este estudio

La figura 17 muestra la cantidad de encuestas realizadas en las diferentes sub regiones del departamento de Nariño, esto se realizo proporcionalmente al tamaño de la población presente

en cada sub region. Como se puede observar las subregiones con mayor número de encuestas son la Región la cordillera; Telembí y Pacifico sur

### 3.4.2 Análisis bivariado

El análisis bivariado mostró asociación a través de la prueba Chi cuadrado (Anexo 3) entre ubicación y consumo, ya que la mayoría de los hogares se encuentran en vereda ( $p<0.05$ ), la mayoría de las viviendas son “Casa” pero esta característica no tubo asociación con el consumo ( $p=0.5734$ ). El uso es residencial y se encontró asociación con el consumo ( $p<0.05$ ), la mayoría de las casas son de ladrillo, con piso de cemento y gravilla con asociación con la variable consumo ( $p<0.05$ ) y finalmente la mayoría de las casas cuentan con alcantarillado estando asociada con el consumo ( $p<0.05$ ).

La base teórica para asociar variables se describe en el anexo 3.

La variable de respuesta “Consumo” se categorizo en los valores, 0 kWh mes; (0-40] Kwh mes; (40-90] kWh mes y >90 kWh mes, teniendo en cuenta que el consumo eficiente es 40 kWh mes y que el consumo mínimo vital es 90 kWh mes (PERS Nariño).

**Tabla 4.**  
**Consumo de energía según características del hogar.**

Característica	Consumo				<i>p</i>
	0 kWh	(0-40] kWh	(40-90] kWh	>90 kWh	
Ubicación					
<b>cabecera municipal</b>	137 (30.0%)	46 (10.0%)	123 (30.0%)	114 (30.0%)	<0.05
<b>Caserío</b>	95 (40.0%)	24 (10.0%)	54 (20.0%)	84 (30.0%)	
<b>centro poblado</b>	6 (70.0%)	1 (10.0%)	2 (20.0%)	0 (0.0%)	
<b>Corregimiento</b>	10 (90.0%)	0 (0.0%)	1 (10.0%)	0 (0.0%)	
<b>inspección de policía</b>	2 (40.0%)	1 (20.0%)	1 (20.0%)	1 (20.0%)	
<b>Vereda</b>	760 (40.0%)	313 (20.0%)	307 (20.0%)	330 (20.0%)	
Tipo de vivienda					
<b>Apartamento</b>	2 (20.0%)	0 (0.0%)	2 (20.0%)	5 (60.0%)	0.5734
<b>Casa</b>	1002 (40.0%)	382 (20.0%)	484 (20.0%)	521 (20.0%)	
<b>Otro</b>	2 (40.0%)	1 (20.0%)	1 (20.0%)	1 (20.0%)	

<b>Sin definir</b>	4 (40.0%)	2 (20.0%)	1 (10.0%)	2 (20.0%)	
Uso					
<b>Residencial</b>	832 (40.0%)	280 (10.0%)	375 (20.0%)	403 (20.0%)	<0.05
<b>Residencial_comercial</b>	63 (40.0%)	18 (10.0%)	44 (30.0%)	48 (30.0%)	
<b>Residencial_cultivos</b>	109 (30.0%)	84 (20.0%)	68 (20.0%)	75 (20.0%)	
<b>Sin especificar</b>	6 (50.0%)	3 (20.0%)	1 (10.0%)	3 (20.0%)	
Pared					
<b>0</b>	0 (0.0%)	0 (0.0%)	1 (50.0%)	1 (50.0%)	-
<b>Guadua</b>	1 (20.0%)	1 (20.0%)	1 (20.0%)	1 (20.0%)	
<b>Ladrillo</b>	319 (20.0%)	384 (30.0%)	453 (30.0%)	204 (20.0%)	
<b>Madera</b>	59 (10.0%)	44 (10.0%)	404 (70.0%)	52 (10.0%)	
<b>Prefabricado</b>	1 (20.0%)	0 (0.0%)	1 (20.0%)	2 (50.0%)	
<b>Sin paredes</b>	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)	
<b>Tapia_bareque</b>	108 (20.0%)	100 (20.0%)	148 (30.0%)	125 (30.0%)	
<b>Zinc_carton</b>	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)	
Piso					
<b>0</b>	1 (50.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)	<0.05
<b>Baldosa,tableta,ladrillo</b>	171 (30.0%)	50 (10.0%)	141 (30.0%)	183 (30.0%)	
<b>Cemento,gravilla</b>	394 (30.0%)	227 (20.0%)	266 (20.0%)	255 (20.0%)	
<b>Madera</b>	340 (70.0%)	39 (10.0%)	44 (10.0%)	48 (10.0%)	
<b>Otro</b>	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
<b>Tierra, arena</b>	103 (40.0%)	68 (30.0%)	37 (10.0%)	43 (20.0%)	
Alcantarillado					
<b>0</b>	8 (30.0%)	7 (30.0%)	3 (10.0%)	6 (20.0%)	<0.05
<b>No</b>	748 (50.0%)	291 (20.0%)	289 (20.0%)	312 (20.0%)	
<b>Si</b>	254 (30.0%)	87 (10.0%)	196 (30.0%)	211 (30.0%)	

Fuente. Este estudio

De acuerdo a la caracterización del consumo por dispositivos de comunicación se encontró que la mayoría de los hogares con mayores consumos no tienen teléfono fijo ( $p=0.015$ ), si tienen celular ( $p<0.05$ ). En contraste, la mayoría de los hogares con los consumos más bajos no tienen dispositivos de radio comunicación ( $p=0.018$ ) y no tienen internet ( $p<0.05$ ).



**Tabla 5.**  
**Consumo de energía según dispositivos de comunicación.**

Característica	Consumo				<i>p</i>
	0 kWh	(0-40] kWh	(40-90] kWh	>90 kWh	
<b>Teléfono</b>					
<b>No</b>	999 (40.0%)	384 (20.0%)	480 (20.0%)	515 (20.0%)	0.015
<b>Si</b>	11 (30.0%)	1 (0.0%)	8 (20.0%)	14 (40.0%)	
<b>Celular</b>					
<b>No</b>	254 (60.0%)	69 (20.0%)	43 (10.0%)	52 (10.0%)	<0.05
<b>Si</b>	756 (40.0%)	316 (20.0%)	445 (20.0%)	477 (20.0%)	
<b>Radio comunicación</b>					
<b>No</b>	999 (40.0%)	385 (20.0%)	487 (20.0%)	528 (20.0%)	0.018
<b>Si</b>	11 (80.0%)	0 (0.0%)	1 (10.0%)	1 (10.0%)	
<b>Internet</b>					
<b>No</b>	984 (40.0%)	382 (20.0%)	473 (20.0%)	495 (20.0%)	<0.05
<b>Si</b>	26 (30.0%)	3 (0.0%)	15 (20.0%)	34 (40.0%)	

Fuente. Este estudio

Los resultados obtenidos del cruce entre características de conexión y electrodomésticos por consumo mostraron que los consumos más bajos estuvieron en hogares sin interconexión. Los consumos más altos fueron en hogares con medidor ( $p<0.05$ ) y que tienen recibo ( $p<0.05$ ), fuente de iluminación lámpara ( $p<0.05$ ), es decir, estas variables son importantes para explicar el consumo. Además, los hogares que tienen refrigerador presentaron los consumos más altos ( $p<0.05$ ). Finalmente, los hogares con aire acondicionado y calefacción presentaron asociación

con el consumo, pero la mayoría de los hogares no cuentan con estos dispositivos por lo tanto es una asociación negativa, es decir, que no es apropiado utilizarlas para explicar el consumo.

**Tabla 6.**

**Consumo de energía según características de conexión y electrodomésticos.**

Característica	Consumo				<i>p</i>
	0 kWh	(0-40] kWh	(40-90] kWh	>90 kWh	
Energía eléctrica					
<b>no</b>	42 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	-
<b>si interconexión</b>	707 (30.0%)	385 (20.0%)	488 (20.0%)	529 (30.0%)	
<b>si planta compartida</b>	76 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
<b>si planta municipal</b>	154 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
<b>si planta propia</b>	31 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Medidor					
<b>No</b>	197 (70.0%)	5 (0.0%)	46 (20.0%)	38 (10.0%)	<0.05
<b>Si</b>	547 (30.0%)	364 (20.0%)	428 (20.0%)	473 (30.0%)	
<b>Sin especificar</b>	266 (80.0%)	16 (10.0%)	14 (0.0%)	18 (10.0%)	
Recibo					
<b>No</b>	972 (90.0%)	28 (0.0%)	31 (0.0%)	20 (0.0%)	<0.05
<b>Si</b>	38 (0.0%)	357 (30.0%)	457 (30.0%)	509 (40.0%)	
Fuente de iluminación					
<b>Energía eléctrica</b>	970 (40.0%)	385 (20.0%)	488 (20.0%)	529 (20.0%)	<0.05
<b>Lámpara de petróleo</b>	40 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Refrigerador					
<b>No</b>	623 (50.0%)	296 (20.0%)	189 (10.0%)	183 (10.0%)	<0.05
<b>Si</b>	387 (30.0%)	89 (10.0%)	299 (30.0%)	346 (30.0%)	
Aire acondicionado					
<b>No</b>	981 (40.0%)	373 (20.0%)	458 (20.0%)	498 (20.0%)	0.003
<b>Si</b>	29 (30.0%)	12 (10.0%)	30 (30.0%)	31 (30.0%)	
Calefacción					
<b>No</b>	984 (40.0%)	377 (20.0%)	456 (20.0%)	493 (20.0%)	<0.05
<b>Si</b>	26 (30.0%)	8 (10.0%)	32 (30.0%)	36 (40.0%)	

Fuente. Este estudio

De acuerdo al análisis bivariado entre el consumo y las características de la forma en la que cocina. Se pudo establecer que la mayoría de los hogares tienen un cuarto exclusivo para cocinar y fueron los que presentaron los consumos más bajos ( $p < 0.05$ ). La mayoría de los hogares concina con gas y leña ( $p < 0.05$ ). Finalmente, la mayoría de hogares gastan muy poca leña por mes y son los hogares con mayor consumo de energía ( $p < 0.05$ ).

**Tabla 7.**

***Consumo de energía según características de la forma de cocinar.***

Característica	Consumo				<i>p</i>
	0 kWh	(0-40] kWh	(40-90] kWh	>90 kWh	
Lugar de cocina					
<b>0</b>	10 (30.0%)	5 (20.0%)	5 (20.0%)	11 (40.0%)	<0.05
<b>Al aire libre</b>	50 (60.0%)	13 (10.0%)	10 (10.0%)	14 (20.0%)	
<b>Cuarto exclusivo cocinar</b>	902 (40.0%)	361 (20.0%)	468 (20.0%)	498 (20.0%)	
<b>Dormitorio</b>	45 (80.0%)	4 (10.0%)	5 (10.0%)	4 (10.0%)	
<b>Ninguna parte</b>	3 (40.0%)	2 (30.0%)	0 (0.0%)	2 (30.0%)	
Combustible para cocinar					
<b>0</b>	2 (30.0%)	1 (20.0%)	0 (0.0%)	3 (50.0%)	<0.05
<b>Carbón vegetal</b>	3 (40.0%)	3 (40.0%)	1 (10.0%)	0 (0.0%)	
<b>Energía eléctrica</b>	20 (50.0%)	2 (10.0%)	9 (20.0%)	7 (20.0%)	
<b>Gas propano</b>	489 (40.0%)	124 (10.0%)	255 (20.0%)	289 (20.0%)	
<b>Leña autoapropiada</b>	405 (40.0%)	216 (20.0%)	169 (20.0%)	168 (20.0%)	
<b>Leña comprada</b>	90 (40.0%)	38 (20.0%)	53 (20.0%)	62 (30.0%)	
<b>Otro</b>	1 (30.0%)	1 (30.0%)	1 (30.0%)	0 (0.0%)	
Leña Kg/Mes					
<b>0 Kg</b>	515 (40.0%)	131 (10.0%)	266 (20.0%)	299 (20.0%)	<0.05
<b>(0-150] Kg</b>	27 (30.0%)	23 (30.0%)	19 (20.0%)	14 (20.0%)	
<b>(150-400] Kg</b>	186 (40.0%)	112 (20.0%)	104 (20.0%)	104 (20.0%)	
<b>&gt;400 Kg</b>	272 (50.0%)	116 (20.0%)	95 (20.0%)	109 (20.0%)	
Aparatos eléctricos					
<b>0KW</b>	115 (70.0%)	26 (20.0%)	8 (10.0%)	10 (10.0%)	<0.05
<b>(0KW-10KW]</b>	402 (40.0%)	229 (20.0%)	232 (20.0%)	216 (20.0%)	
<b>(10KW-20KW]</b>	240 (40.0%)	73 (10.0%)	129 (20.0%)	129 (20.0%)	
<b>(20KW-50KW]</b>	198 (40.0%)	44 (10.0%)	99 (20.0%)	119 (30.0%)	
<b>&gt;50</b>	55 (40.0%)	13 (10.0%)	20 (10.0%)	55 (40.0%)	

Fuente. Este estudio

### 3.4.3 Modelación del consumo

Para modelar el consumo se recodificó la variable consumo promedio en (0: Consumos entre 0 y 40 kWh y 1: Consumos > 40 kWh), teniendo en cuenta el consumo eficiente, y se propuso un modelo de regresión lineal generalizado con respuesta binomial y familia logit que se describe en el anexo 4, para ajustar el consumo en función de la ubicación, uso de la vivienda, si usa radio comunicación, energía eléctrica, recibo, fuente de iluminación, si tiene refrigerador, lugar de cocina y consumo de leña en kg/mes. El modelo se ajustó con la función *glm* (modelo lineal generalizado) del paquete *stats* en el programa estadístico R versión 3.6. El modelo de regresión logística, el método de estimación y la prueba de hipótesis se describen en el anexo 4. A continuación se presenta la salida del primer ajuste. Los resultados mostraron que las variables que explican el consumo fueron la ubicación, radio comunicación, recibo, refrigerador y el consumo de leña, es decir que son variables con las cuales se puede predecir el consumo de un hogar.

**Tabla 8.**  
**Estimación de los parámetros del modelo inicial.**

Coeficientes	Estimación	Error estándar	Estadístico	p
<b>(Intercept)</b>	-17.931	2419.823	-0.007	0.994
<b>UBICACIONcaserío</b>	-0.018	0.233	-0.075	0.940
<b>UBICACIONcentro poblado</b>	-1.653	1.220	-1.356	0.175
<b>UBICACIONcorregimiento</b>	-0.656	1.738	-0.377	0.706
<b>UBICACIONinspección de policía</b>	-0.071	1.499	-0.047	0.962
<b>UBICACIONvereda</b>	-0.676	0.179	-3.780	0.000
<b>USOresidencial_comercial</b>	-0.101	0.228	-0.444	0.657
<b>USOresidencial_cultivos</b>	0.172	0.169	1.013	0.311
<b>USOsin especificar</b>	-0.563	0.789	-0.714	0.475
<b>RADIO_COMUNICACIONsi</b>	2.395	1.314	1.823	0.068
<b>ENERGIA_ELECTRICAsi interconexión</b>	15.207	2419.823	0.006	0.995
<b>ENERGIA_ELECTRICAsi planta compartida</b>	-2.678	2505.661	-0.001	0.999
<b>ENERGIA_ELECTRICAsi planta municipal</b>	-0.321	2444.327	0.000	1.000
<b>ENERGIA_ELECTRICAsi planta propia</b>	0.064	2638.495	0.000	1.000
<b>RECIBOsi</b>	3.856	0.173	22.331	< 2e-16
<b>FUENTE_ILUMINACIONlampara de petroleo</b>	0.190	2479.985	0.000	1.000
<b>REFRIGERADORsi</b>	1.395	0.134	10.374	< 2e-16
<b>LUGAR_COCINAal aire libre</b>	-0.722	0.650	-1.111	0.267
<b>LUGAR_COCINAcuarto exclusivo cocinar</b>	-0.680	0.547	-1.243	0.214
<b>LUGAR_COCINAdormitorio</b>	-1.228	0.776	-1.582	0.114
<b>LUGAR_COCINAninguna parte</b>	-1.266	1.154	-1.098	0.272
<b>LENA_Kg_MES(150-400] Kg</b>	0.566	0.305	1.854	0.064
<b>LENA_Kg_MES&gt;400 Kg</b>	0.217	0.302	0.717	0.473
<b>LENA_Kg_MES0 Kg</b>	0.243	0.301	0.808	0.419

Fuente. Este estudio

El segundo modelo ajustado sugiere que las variables que efectivamente sirven para explicar el consumo fueron la ubicación ( $p < 0.05$ ) ya que discrimina muy bien los hogares con consumos altos y bajos ya sea si es cabecera municipal o vereda. La segunda variable importante para explicar el consumo fue si el hogar tiene recibo ( $p < 0.05$ ) ya que los consumos más altos estuvieron en los

hogares con recibo, es lógico puesto que permite cuantificar adecuadamente el consumo. La tercera variable es si el hogar tiene refrigerador, ya que está asociada a los consumos más altos, por lo tanto, dentro del grupo de electrodomésticos fue el que mejor discriminó el consumo y finalmente la variable consumo de leña sirve bastante para discriminar el consumo ( $p=0.063$ ) con una significancia del 10%.

**Tabla 9.**

*Estimación de los parámetros del modelo final.*

Coeficientes	Estimación	Error estándar	Estadístico	p
<b>(Intercept)</b>	-3.55616	0.35158	-10.115	< 2e-16
<b>UBICACIONcaserío</b>	0.02097	0.23895	0.088	0.9301
<b>UBICACIONcentro poblado</b>	-1.59939	1.12347	-1.424	0.1546
<b>UBICACIONcorregimiento</b>	-1.014	1.54426	-0.657	0.5114
<b>UBICACIONinspección de policía</b>	-0.02447	1.55279	-0.016	0.9874
<b>UBICACIONvereda</b>	-0.71456	0.17801	-4.014	5.97E-05
<b>RADIO_COMUNICACIONsi</b>	0.21308	1.03773	0.205	0.8373
<b>RECIBOsi</b>	4.0469	0.16721	24.202	< 2e-16
<b>REFRIGERADORsi</b>	1.41358	0.13205	10.705	< 2e-16
<b>LENA_Kg_MES(150-400] Kg</b>	0.56184	0.30267	1.856	0.0634
<b>LENA_Kg_MES&gt;400 Kg</b>	0.21264	0.30089	0.707	0.4797
<b>LENA_Kg_MES0 Kg</b>	0.22593	0.29811	0.758	0.4485

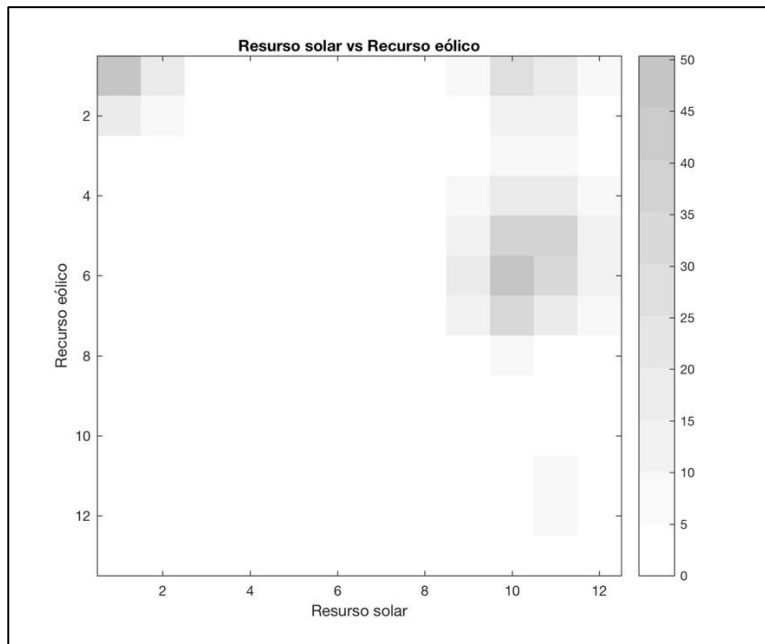
Fuente. Este estudio

### 3.4.4 Análisis correlacional de las variables numéricas presentes en la base de datos

inicial.

Se realizó un análisis correlacional con las variables continuas en el programa MATLAB. Las correlaciones más importantes, resultado del análisis mediante el histograma dos dimensiones permiten observar, por ejemplo, las localidades con recurso solar bajo presentan baja afluencia de viento bajo correspondiendo a algunos sectores cercanos a la costa pacífica y amazónica (figura 18). Por otra parte, en sitios donde se cuenta con recurso solar alto la velocidad de viento es

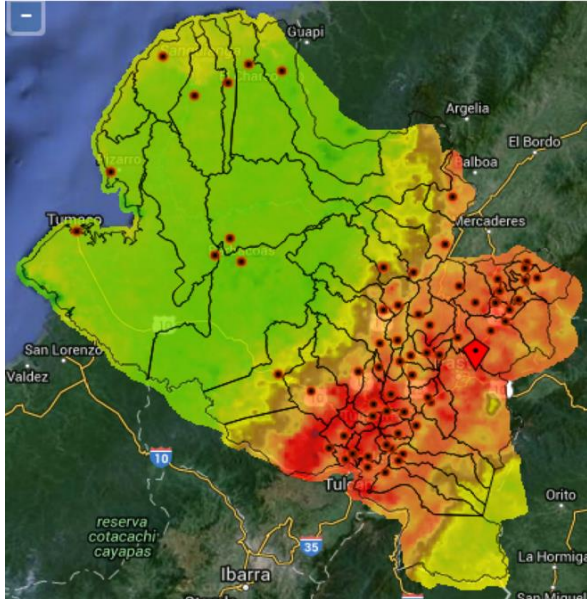
considerablemente alto coincidiendo con los sectores del departamento sobre la Cordillera de los Andes hacia la Amazonía.



**Figura 18. Relación entre recurso solar ( $\text{kW}/\text{m}^2$ ) y velocidad del viento ( $\text{m}/\text{s}^2$ )**

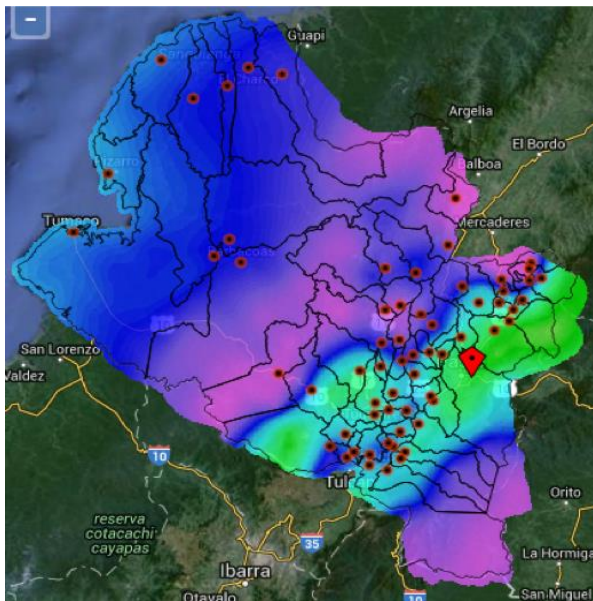
Fuente. Este estudio

Esto se puede observar en los mapas de caracterización del recurso solar y eólico del departamento de Nariño obtenidos en ALTERNAR:



**Figura 19. Mapa del recurso solar para el Departamento de Nariño**

Fuente. Este estudio

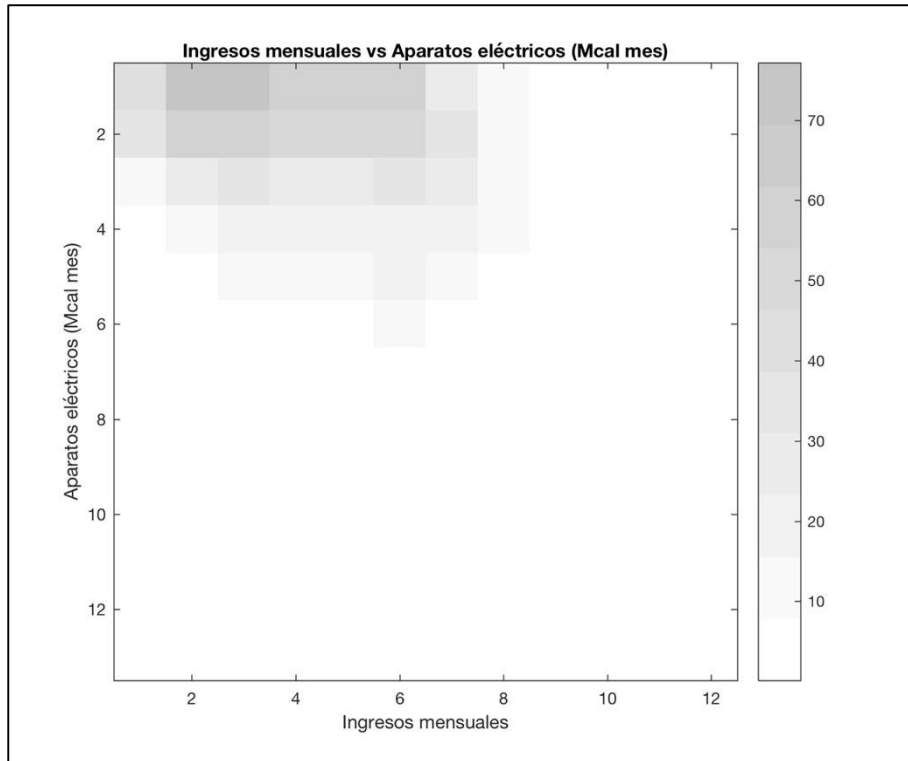


**Figura 20. Mapa del recurso eólico para el Departamento de Nariño**

Fuente. Este estudio

Partiendo de la información obtenida, se observa que un gran número de hogares con ingresos bajos presentan un consumo bajo de energía en aparatos eléctricos menor o igual a 2 Mcal al mes.

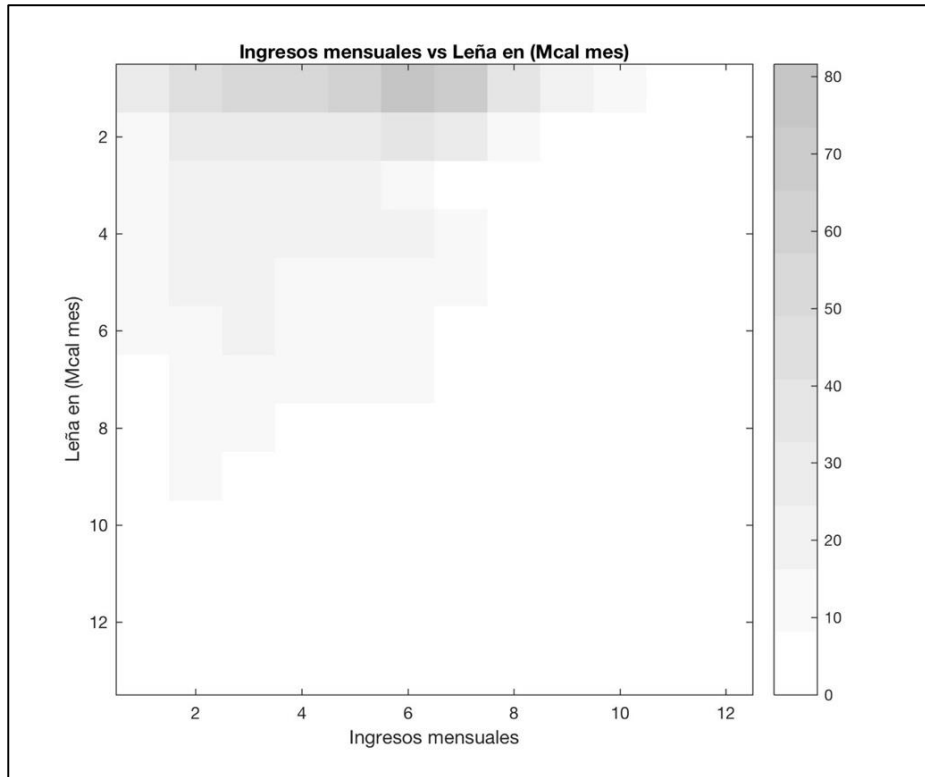




**Figura 21. Relación entre consumo de energía en aparatos eléctricos e ingresos mensuales en los hogares**

Fuente. Este estudio

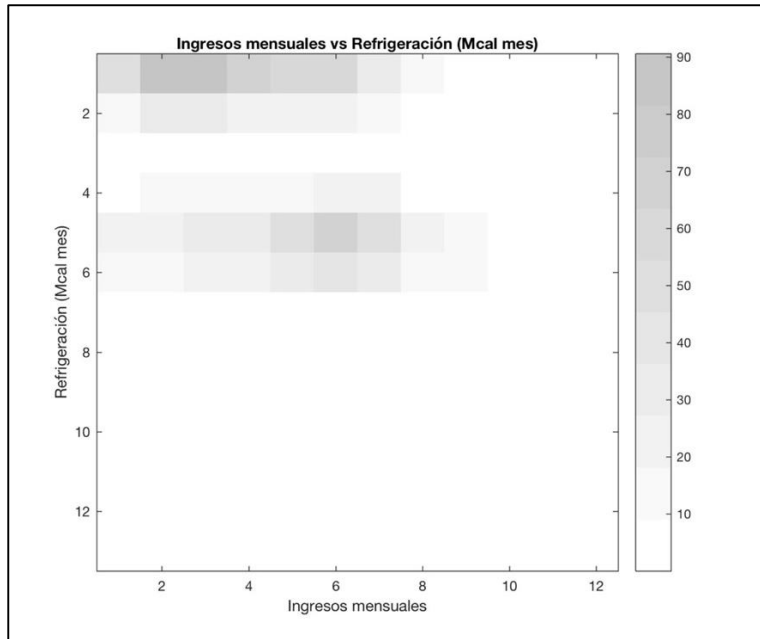
A partir de la figura 22, se puede concluir que el consumo de leña en el departamento de Nariño se orienta hacia hogares con ingreso bajos y medio.



**Figura 22. Relación entre el consumo de leña e ingresos mensuales de los hogares**

Fuente. Este estudio

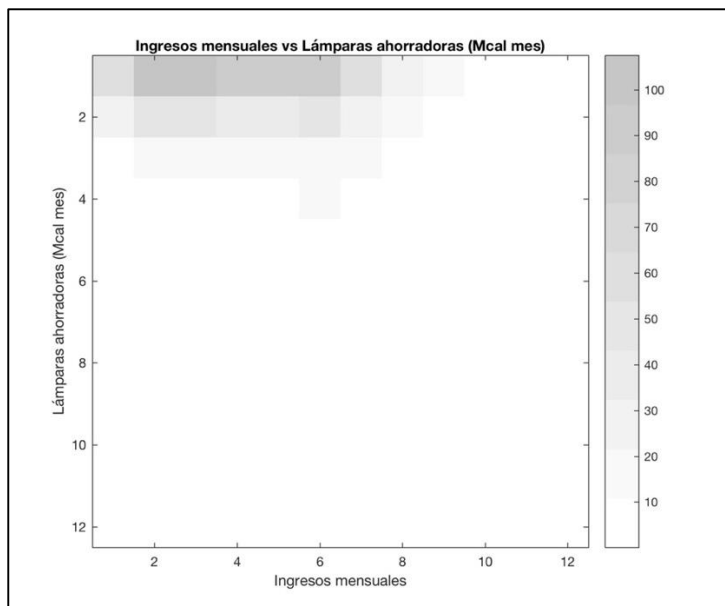
Un consumo bajo de energía para refrigeración menor a 2 Megacalorías al mes es mayoritariamente frecuente para los hogares. También se presenta un consumo medio entre 4 y 6 Megacalorías para hogares con ingreso medio (figura 23).



**Figura 23. Relación entre ingresos mensuales y uso de energía en refrigeración**

Fuente. Este estudio

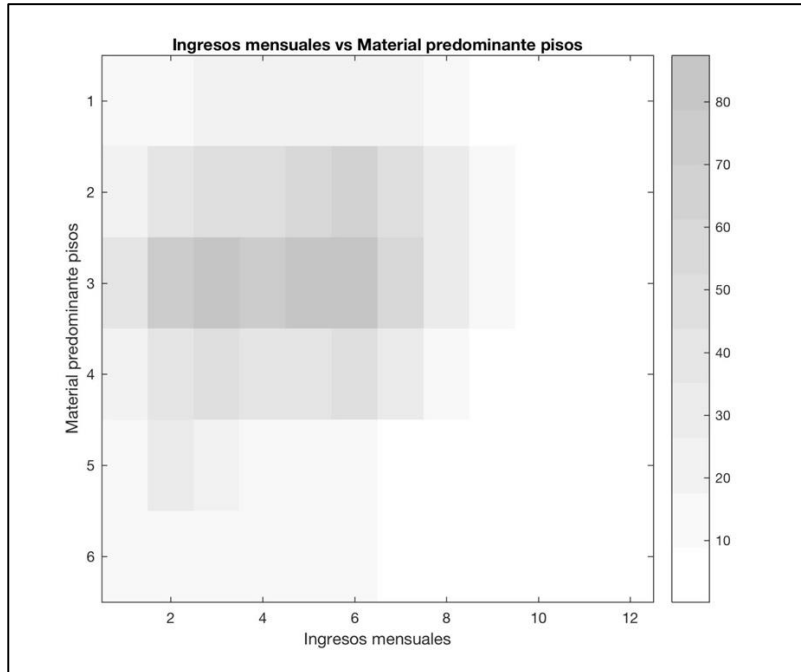
El uso de lámparas ahorradoras es bajo, correspondiendo hogares con consumos mayoritariamente menores a 2 Megacalorías e ingresos medios y bajos (figura 9).



**Figura 24. Relación entre el uso de lámparas ahorradoras e ingresos mensuales en el hogar**

Fuente. Este estudio

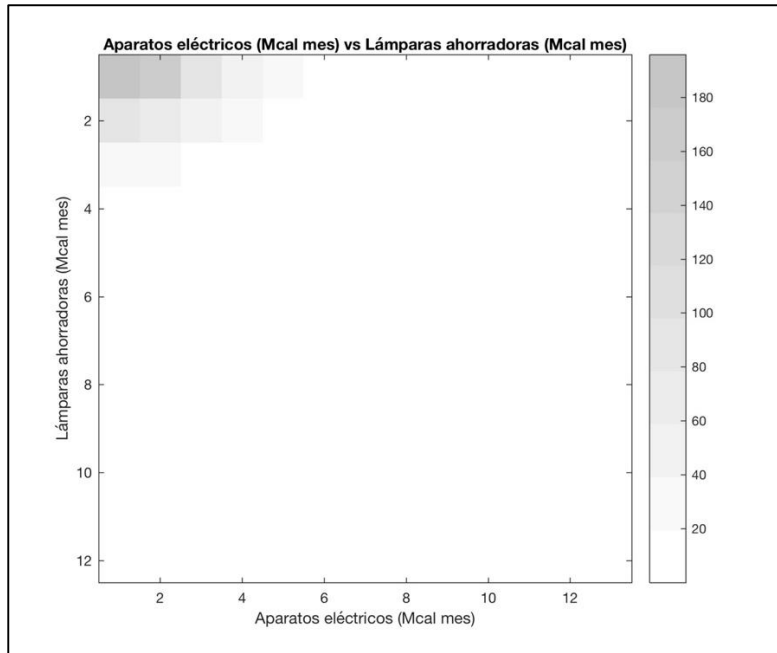
Aquellos hogares con ingresos mensuales superiores a \$100.000 y menores a \$750.000 prevalecen en viviendas con pisos de materiales madera burda, tabla, tablón u otro vegetal (figura 10).



**Figura 25. Relación entre el material predominante en pisos e ingresos mensuales en el hogar**

Fuente. Este estudio

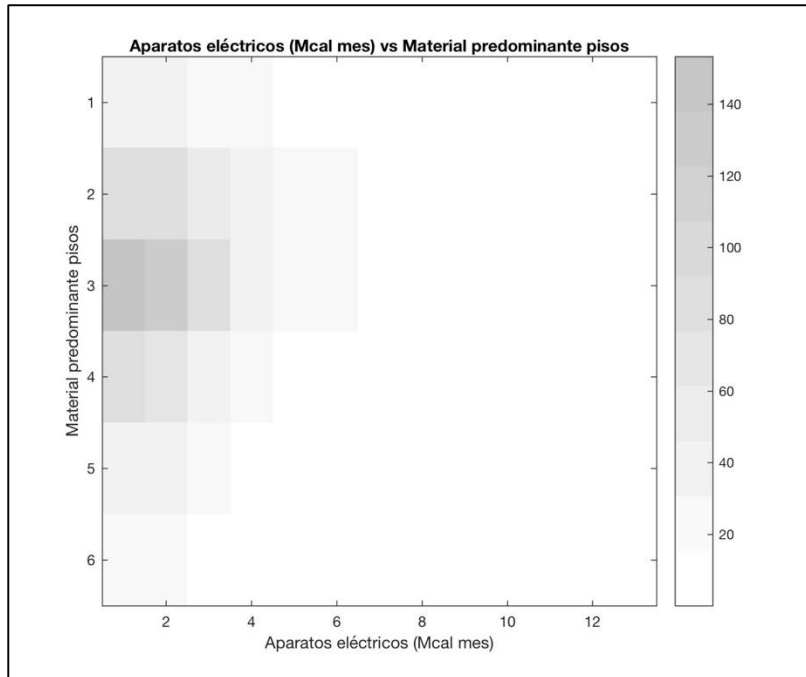
El consumo de aparatos eléctricos se relaciona de manera fuerte con el uso de lámparas ahorradoras para iluminación.



**Figura 26. Relación entre el uso de lámparas ahorradoras y uso de aparatos eléctricos**

Fuente. Este estudio

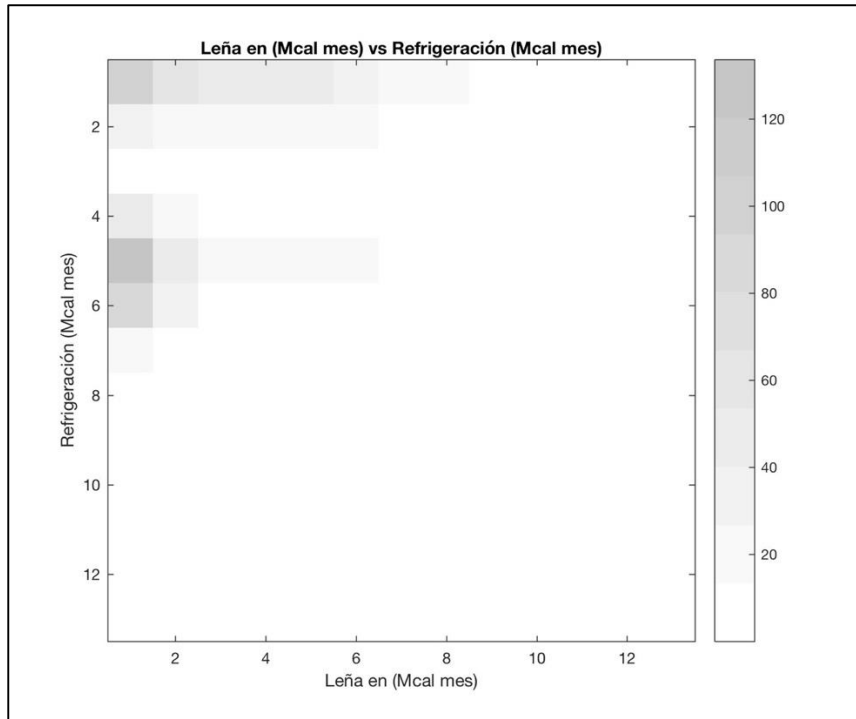
Resulta interesante la relación existente entre el consumo de energía en aparatos eléctricos y el tipo material del piso de la vivienda que predomina. El tipo predominante de piso madera burda, tabla, tablón u otro vegetal es donde mayoritariamente se consume energía (figura 27).



**Figura 27. Relación entre el uso de aparatos eléctricos y el material predominante en los pisos de la vivienda**

Fuente. Este estudio

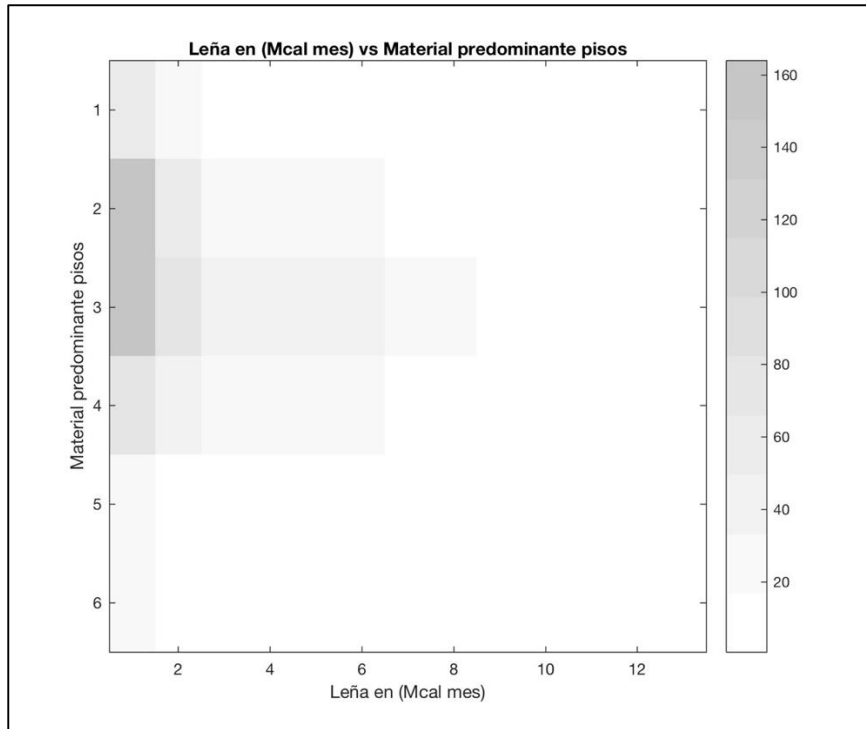
Muchos de los hogares que consumen cerca de una Megacaloría en leña al mes hacen uso de energía en refrigeración con cerca de 5 Megacalorías al mes (Figura 28).



**Figura 28. Relación entre el uso leña y refrigeración en el hogar**

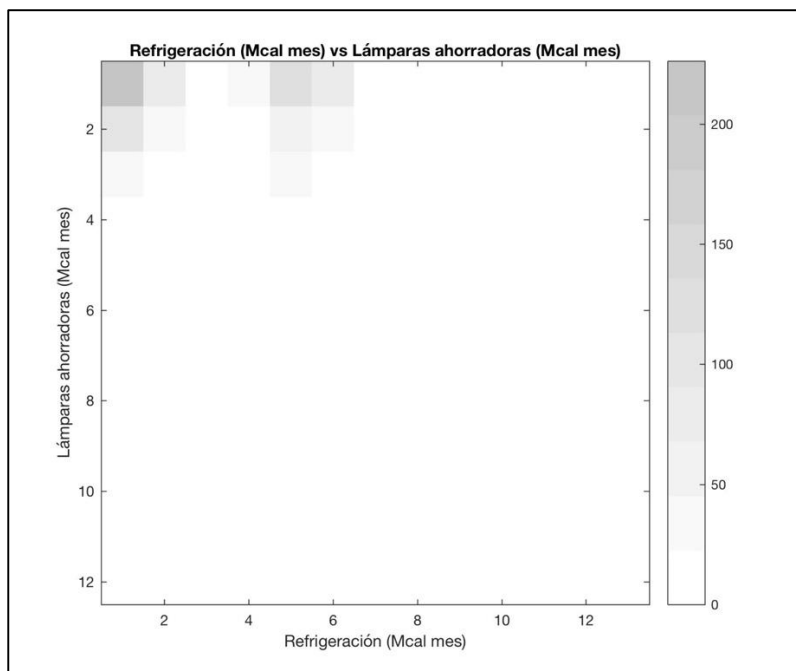
Fuente. Este estudio

Es habitual que, en hogares con pisos de madera burda, tabla, tablón, otro vegetal y cemento o gravilla se consuma menos que una Megacaloría por uso de leña (Figura 29).



**Figura 29. Relación entre el consumo de leña vs material predominante pisos**

Fuente. Este estudio

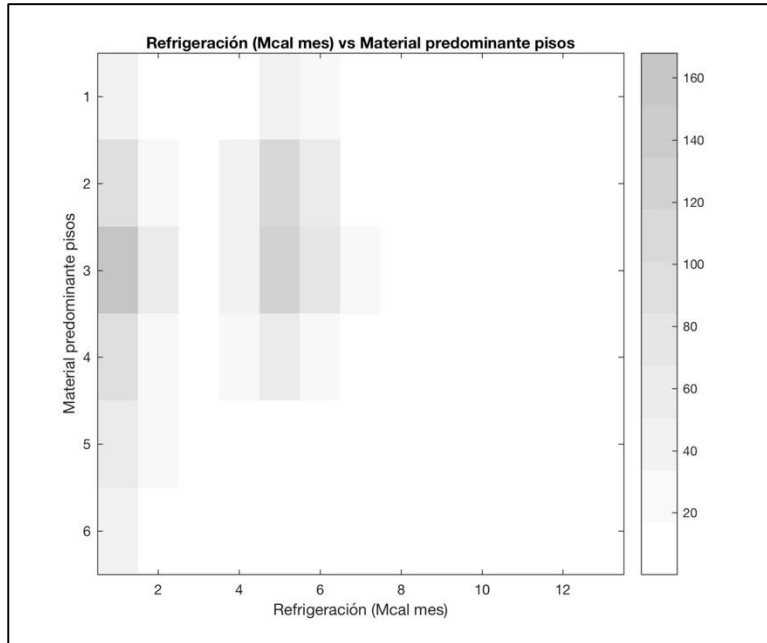


**Figura 30. Relación entre refrigeración y lámparas ahorradoras**

Fuente. Este estudio



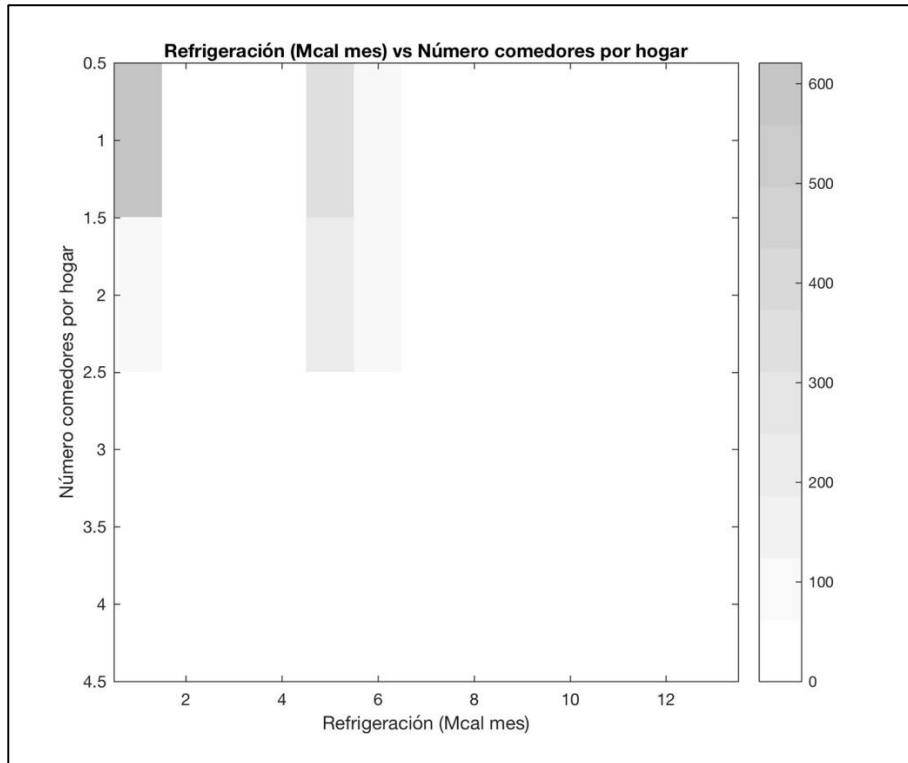
Los consumidores más frecuentemente encontrados para refrigeración son aquellos hogares que tiene pisos de madera burda, tabla, tablón u otro vegetal (Figura31).



**Figura 31. Relación entre refrigeración vs el material predominante en pisos**

Fuente. Este estudio

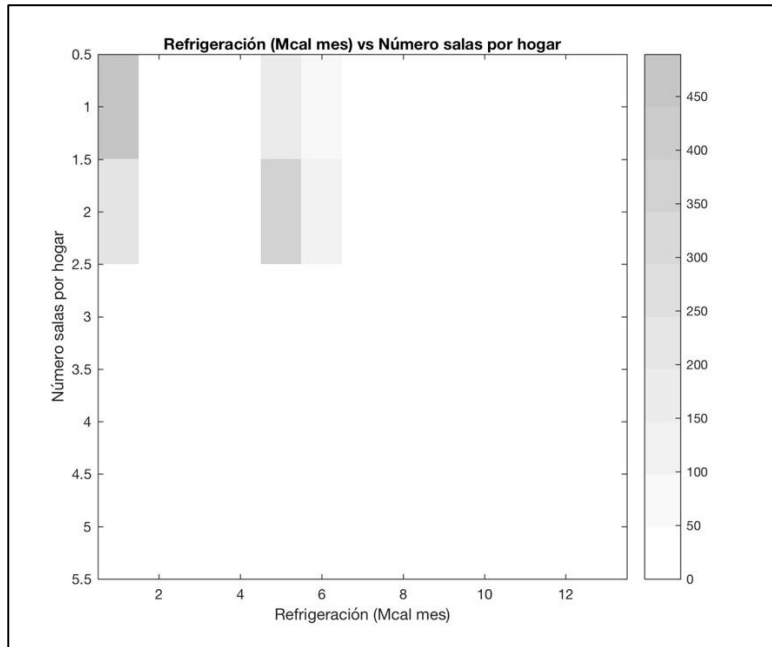
Se consumen habitualmente 1 o 5 Megacalorías para uso de refrigeración. El consumo menor es para aquellos hogares con un solo comedor.



**Figura 32. Relación entre refrigeración vs número de comedores**

Fuente. Este estudio

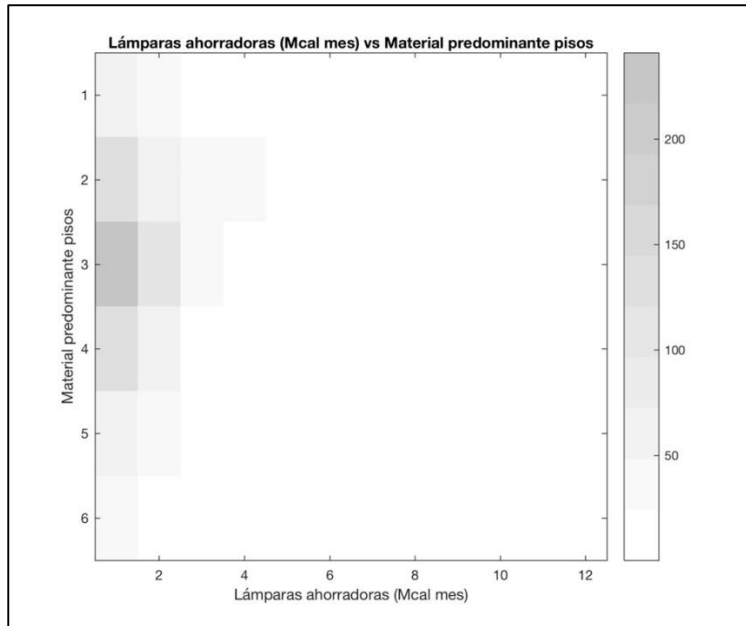
Los hogares con una sola sala presentan un bajo consumo en refrigeración contrastando con los hogares que tienen dos salas donde el consumo más habitual está alrededor de 5 Megacalorías al mes.



**Figura 33. Relación entre refrigeración vs número de salas por hogar**

Fuente. Este estudio

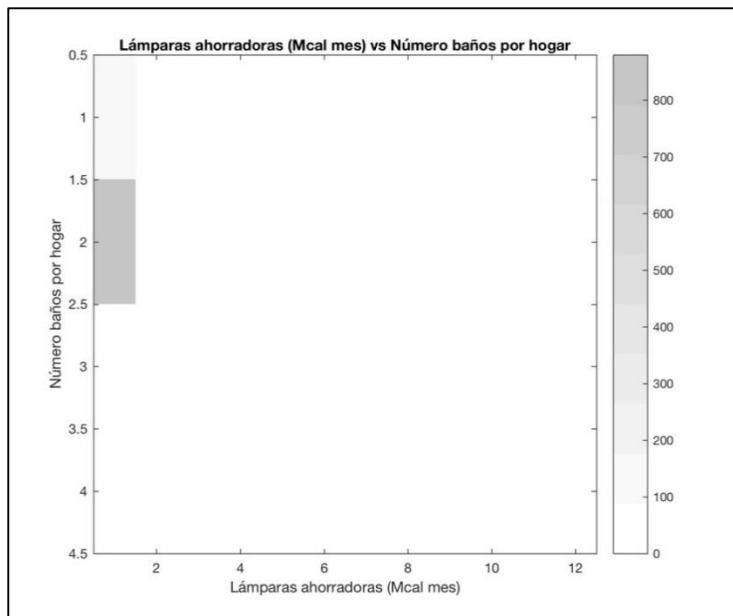
El uso de lámparas ahorradoras es bastante bajo, se usa muy frecuentemente una en hogares que tienen pisos de madera burda, tabla, tablón u otro vegetal,



**Figura 34. Lámparas ahorradoras vs material predominante en pisos**

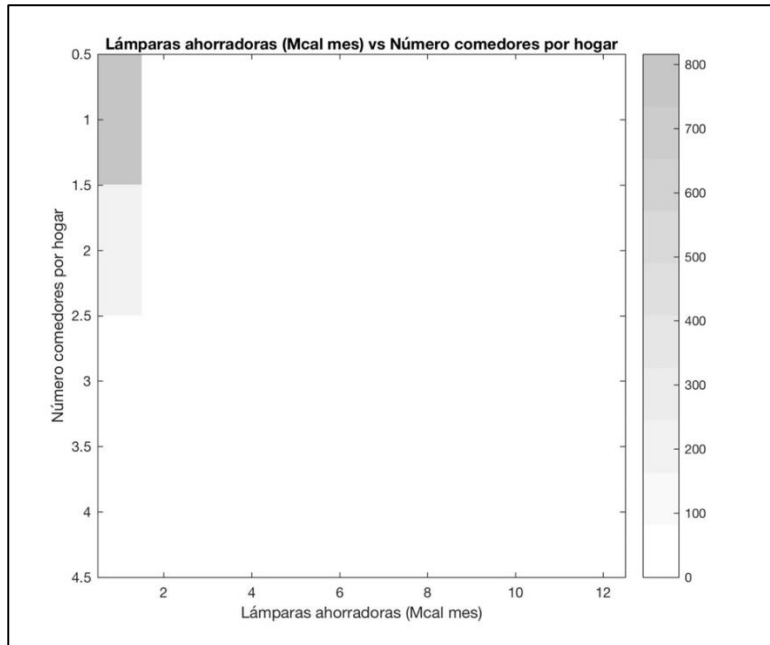
Fuente. Este estudio

Aquellos hogares que cuentan con dos baños y un comedor por hogar hacen uso de una bombilla ahorradora mayoritariamente (Figura 35 y 36)



**Figura 35. Lámparas ahorradoras vs número de baños por hogar**

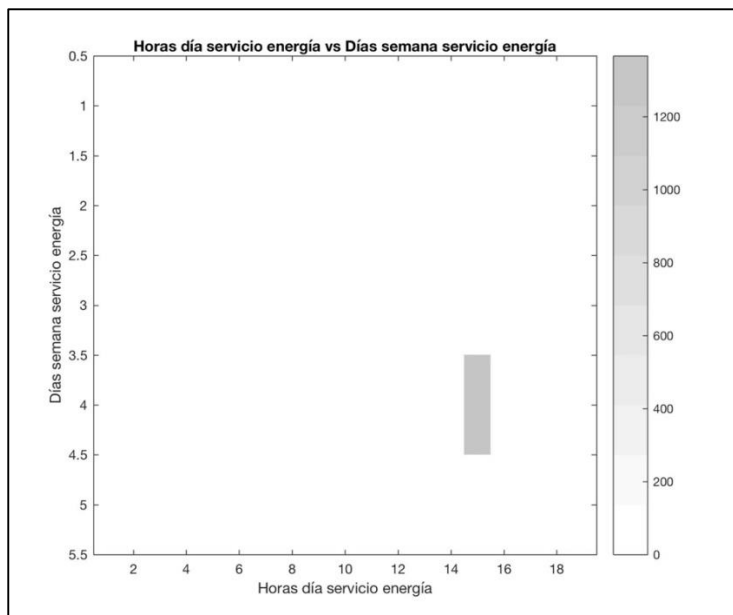
Fuente. Este estudio



**Figura 36. Lámparas ahorradoras vs número de comedores por hogar**

Fuente. Este estudio

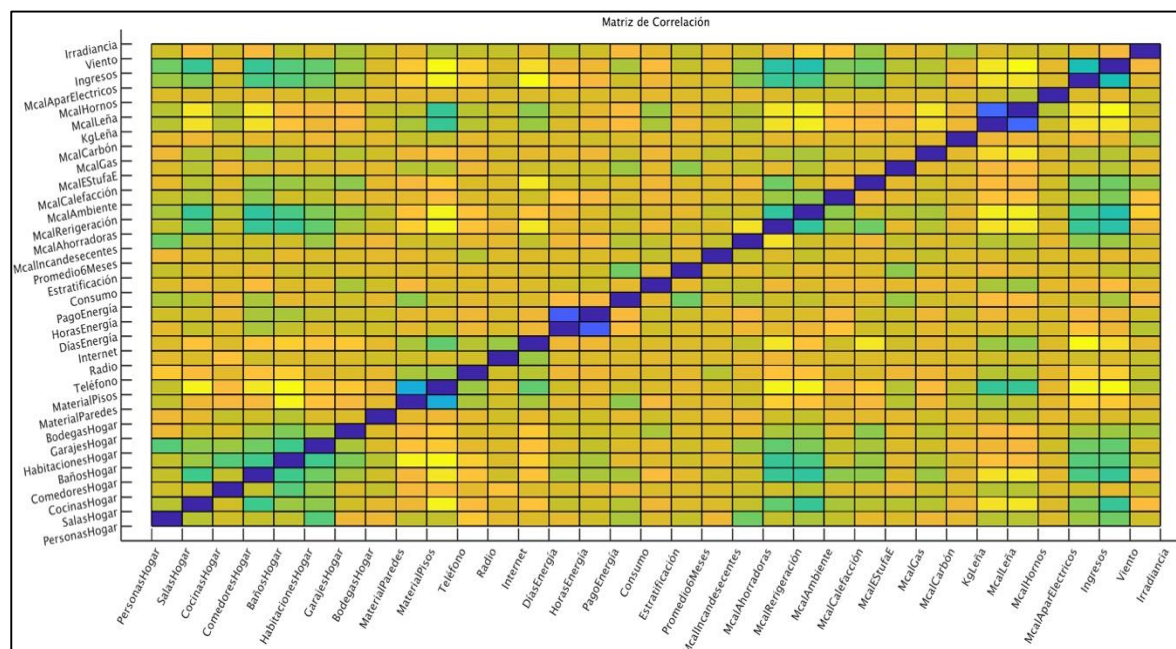
Aquellos hogares que manifiestan contar con 4 días de servicio de energía también indican que en la prestación del servicio de energía, reciben 15 horas de servicio al día.



**Figura 37. Horas al día servicio de energía vs días en semana de servicio eléctrico**

Fuente. Este estudio

## Matriz de Correlación

**Figura 38. Matriz de correlación**

Fuente. Este estudio

Las correlaciones más importantes se muestran en la tabla 10.

**Tabla 40.****Correlaciones más relevantes-**

Resultados de la investigación	Evidencias
El nivel de ingresos de los hogares se encuentra correlacionado con el uso de energía para lámparas ahorradoras, aparatos eléctricos y el consumo de leña.	Figuras: 22, 24 y 26
El consumo eficiente de energía en aparatos eléctricos presenta una fuerte correlación con el bajo consumo de los hogares que usan lámparas ahorradoras	Figura 26
Localidades con alta afluencia de viento presentan baja intensidad de brillo solar y horas de sol. Este resultado se evidencia particularmente para las cimas de las cordilleras andinas de acuerdo a los mapas de recurso solar y eólico para Nariño.	Figuras 18, 19 y 20
El consumo de energía empleado en refrigeración presenta dos grupos frecuentes, consumo alto y bajo, esto es más evidente cuando se confronta con las variables número de comedores, material predominante de pisos, uso de lámparas ahorradoras y consumo de leña.	Figuras 28, 30, 31, 32 y 33.

#### 4. Conclusiones

El objetivo de la presente investigación fue identificar las principales variables, que influyen en el consumo energético en las trece sub regiones del departamento de Nariño, lo cual se pudo determinar a partir de los análisis tanto univariado, bivariado y multivariado, que existen variables que explican el consumo eléctrico residencial en las zonas rurales del departamento o de Nariño, estas están asociadas al consumo de leña, condiciones de tenencia de electrodomésticos y ubicación de la residencia, en concordancia a lo establecido por Karatasoua, S. y Santamouris, M. (2019); Subbiah, R., Pal, A., Nordberg, E., Marathe, A., y Marathe, M. (2017) y Swan, L., Ugursal, I. (2009).

Hecho el análisis del estudio metodología de la recolección primaria hecha por el PERS Nariño en el año 2013, se pudieron determinar diferencias significativas entre grupos de consumidores de energía eléctrica y se determinó, que variables como consumo de leña, tenencia de electrodomésticos y ubicación, son las que más aportan información a la variable de respuesta que es consumo de aparatos eléctricos.

En muchos casos cuando la dimensionalidad es muy grande, se hace necesario una reducción de variables, teniendo en cuenta que se descartan las que presentan muchos valores faltantes, variables muy correlacionadas con otras, que resultan de procesos de conversiones, y también en las cuales se utiliza la suma de variables que trabajan con las mismas unidades y en la misma escala.

Asentamientos poblados ubicados en zonas geográficas muy apartadas no cuentan con servicio de interconexión y por tanto utilizan fuentes de energía no renovables como derivados del petróleo para las actividades como cocinar.

En estudios en donde se cuente con una gran cantidad de información, tanto de atributos o campos como de registros, se hace bastante útil, la metodología CRISP DM, la cual posibilita un manejo favorable que permite la obtención de modelos que explican el comportamiento de las variables dentro del mismo. Para el caso de la presente investigación esta metodología fue de bastante utilidad porque permitió resumir los campos de estudio, eliminar redundancias y obtener información relevante a través de un modelo y asociación entre variables.

Se pudo evidenciar que, en sitios del territorio del Departamento de Nariño, donde se presentan alta afluencia de viento, el recurso solar es bajo, este resultado es confirmado por los mapas de recurso solar y eólico desarrollado en el proyecto ALTERNAR. Estas ubicaciones corresponderían a las cimas de las cordilleras andinas con pocas horas e intensidad de brillo solar.

El consumo eficiente de energía en aparatos eléctricos presenta una fuerte correlación con el bajo consumo de los hogares que usan lámparas ahorradoras. Este resultado presenta una orientación importante para las políticas de uso racional y eficiente de energía, donde el sector de iluminación logra un gran impacto en la disminución de uso de energía y por ende un ahorro por esta disminución. Es decir, la introducción de aparatos eléctricos eficientes energéticamente en los hogares rurales nariñenses se hace en la misma dirección que los sistemas de iluminación eficientes.

En concordancia con los resultados obtenidos por los trabajos expuestos en los antecedentes, esta investigación muestra que el nivel de ingresos de los hogares se encuentra correlacionado con el uso de energía para lámparas ahorradoras y aparatos eléctricos. Particularmente, en el análisis



expuesto, encontramos una relación adicional entre el consumo de leña y el nivel de ingresos, que no se expone en los documentos de la revisión bibliográfica.

La categorización de la variable respuesta, consumo energía eléctrica de la vivienda, permitió identificar la demanda de las viviendas rurales en tres grupos: consumo eficiente (0,40kWh], consumo medio (40kWh, 90kWh] por debajo del consumo mínimo vital y consumo alto por encima de los 90kWh.

## **5. Recomendaciones**

Hacer el análisis subregiones, teniendo en cuenta que existen diferencias significativas entre grupos y fuertes evidencias sobre las relaciones entre uso de leña y consumo eléctrico, lo cual en Nariño se acentúa al tener estas diferentes condiciones culturales geográficas y de pisos térmicos

Realizar proyección de aprovechamiento de energías renovables locales, como la eólica y la solar, de acuerdo a las condiciones presentes en cada subregión.

Se sugiere el trabajo utilizando variables de tipo ambiental para determinar qué tan factible y optimo resultaría la implementación de sistemas de aprovechamiento de energía solar y eólica.

Trabajar con otras técnicas estadísticas o de inteligencia artificial que permitan modelar el consumo de energía desde la información base.

Trabajar paralelamente modelos cualitativos y cuantitativos para el análisis de consumo eléctrico.

Realizar investigaciones de consumo eléctrico en pueblos y ciudades y compararlos con el consumo eléctrico rural. Gran encuesta integrada de hogares del DANE 2018.



### Bibliografía

- Luna, V., & Cevallos, L. (2015). ESTUDIO DE LA DEMANDA ELÉCTRICA DIVERSIFICADA EN EDIFICIOS RESIDENCIALES. (*Tesis de ingeniería*). Universidad Politecnica Salesiana Sede Quito, Quito.
- Aydin, E., & Brounen, D. (2019). The impact of policy on residential energy consumption. *Energy* 169, 115-119.
- Hernández Orallo, J., Ramírez Quintana, J., & Ferri Ramírez, C. (2004). *Introduccion a la minería de datos*. Madrid: Pearson educación, S.A.
- Ahmad, J., Larijani H., Emmanuel, R., Mannion, M., and Javed, A. (2017). Energy Demand Prediction Through Novel Random Neural Network Predictor for Large Non-Domestic Buildings. Annual IEEE International Systems Conference.
- Departamento Nacional de Planeación (2017). Energy Demand Situation in Colombia. Recuperado de: <https://www.dnp.gov.co/Crecimiento-Verde/Documents/ejes-tematicos/Energia/MCV%20-%20Energy%20Demand%20Situation%20VF.pdf>
- Huebner, G., Shipworth, D., Hamilton, I., Chalabi, Z. y Oreszczyn, T. (2016). Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. *Applied Energy*, 177, 692–702.
- Jones, R., Fuertes, A., y Lomas K, (2014). The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews*, 43, 901-917.
- Karatasoua, S. y Santamouris, M. (2019). Socio-economic status and residential energy consumption: A latent variable approach. *Energy & Buildings*, 198. 100–105
- Rahman, R., Sajib, N., Rifat, S., Haider, H., Khan, A. (2016). Forecasting the Long Term Energy Demand of Bangladesh Using SPSS from 2011-2040. 3rd International Conference on

Electrical Engineering and Information Communication Technology (ICEEICT).

Stephens, E., Smith, D., y Mahanti, A. (2015). Game Theoretic Model Predictive Control for Distributed Energy Demand-Side Management. *IEEE Smart Grid Trans.*, 6, 3, 1394-1402.

Subbiah, R., Pal, A., Nordberg, E., Marathe, A., y Marathe, M. (2017). Energy Demand Model for Residential Sector : A First Principles Approach. *IEEE Sustainable Energy Trans.*, 8, 3, 1215-1224.

Swan, L., Ugursal, I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13, 8, 1819-1835.

Unidad de Planeación Minero Energética, UPME (2019). Proyección de Demanda de Energía Eléctrica en Colombia, Revisión Febrero de 2019.

## Anexos

### Anexo I. Análisis Bivariado

#### Script:

```
#Librerias
install.packages("missMDA")

rm(list = ls())

library(missMDA)
library(readxl)
library(data.table)
library(ggplot2)
library(dplyr)
library(openxlsx)

library(ade4)
library(FactoMineR)
library(factoextra)

Bd_Analisis <- as.data.frame(read_excel("Bd_Analisis_ACM.xlsx"))

#####
#####
#ANALISIS BIVARIADO
#####
#####

devtools::install_github("stulacy/epitab")

library(epitab)
library(dplyr)
library(knitr)

ruta <- "/Users/johannospina/Dropbox/ARTICULOS_EN_PROCESO/ANALISIS_TESIS_ANDRES_CHECA/Tablas/"

tbl_1 <- Bd_Analisis %>%
  group_by(UBICACION, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1))

tbl_1$export <- with(tbl_1, sprintf("%i (%.1f%%)", n, por*100))
tbl_1 <- dcast(UBICACION~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_1, fill="0 (0.0%)")
write.xlsx(tbl_1, paste0(ruta,"tbl_1.xlsx"), sheetName = "tbl_1",
  col.names = TRUE, row.names = FALSE, append = FALSE)

chisq.test(table(Bd_Analisis$UBICACION, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_2 <- Bd_Analisis %>%
  group_by(TIPO_VIVIENDA, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_2$export <- with(tbl_2, sprintf("%i (%.1f%%)", n, por*100))
tbl_2 <- dcast(TIPO_VIVIENDA~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_2, fill="0 (0.0%)")
write.xlsx(tbl_2, paste0(ruta,"tbl_2.xlsx"), sheetName = "tbl_2",
  col.names = TRUE, row.names = FALSE, append = FALSE)

chisq.test(table(Bd_Analisis$TIPO_VIVIENDA, Bd_Analisis$CONSUMOKwh_AGRUPADO))
```

```

tbl_3 <- Bd_Analisis %>%
  group_by(USO, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_3$export <- with(tbl_3, sprintf("%i (%.1f%%)", n, por*100))
tbl_3 <- dcast(USO~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_3, fill="0 (0.0%)")
write.xlsx(tbl_3, paste0(ruta,"tbl_3.xlsx"), sheetName = "tbl_3",
  col.names = TRUE, row.names = FALSE, append = FALSE)

chisq.test(table(Bd_Analisis$USO, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_4 <- Bd_Analisis %>%
  group_by(PARED, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_4$export <- with(tbl_4, sprintf("%i (%.1f%%)", n, por*100))
tbl_4 <- dcast(PARED~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_4, fill="0 (0.0%)")
write.xlsx(tbl_4, paste0(ruta,"tbl_4.xlsx"), sheetName = "tbl_4",
  col.names = TRUE, row.names = FALSE, append = FALSE)

chisq.test(table(Bd_Analisis$PARED, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_5 <- Bd_Analisis %>%
  group_by(PISO, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_5$export <- with(tbl_5, sprintf("%i (%.1f%%)", n, por*100))
tbl_5 <- dcast(PISO~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_5, fill="0 (0.0%)")
write.xlsx(tbl_5, paste0(ruta,"tbl_5.xlsx"), sheetName = "tbl_5",
  col.names = TRUE, row.names = FALSE, append = FALSE)

chisq.test(table(Bd_Analisis$PISO, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_6 <- Bd_Analisis %>%
  group_by(ALCANTARILADO, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_6$export <- with(tbl_6, sprintf("%i (%.1f%%)", n, por*100))
tbl_6 <- dcast(ALCANTARILADO~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_6, fill="0 (0.0%)")
write.xlsx(tbl_6, paste0(ruta,"tbl_6.xlsx"), sheetName = "tbl_6",
  col.names = TRUE, row.names = FALSE, append = FALSE)

chisq.test(table(Bd_Analisis$ALCANTARILADO, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_7 <- Bd_Analisis %>%
  group_by(TELEFONO, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_7$export <- with(tbl_7, sprintf("%i (%.1f%%)", n, por*100))
tbl_7 <- dcast(TELEFONO~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_7, fill="0 (0.0%)")
write.xlsx(tbl_7, paste0(ruta,"tbl_7.xlsx"), sheetName = "tbl_7",
  col.names = TRUE, row.names = FALSE, append = FALSE)

```

```
chisq.test(table(Bd_Analisis$TELEFONO, Bd_Analisis$CONSUMOKwh_AGRUPADO))
```

```
tbl_8 <- Bd_Analisis %>%
  group_by(CELULAR, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_8$export <- with(tbl_8, sprintf("%i (%.1f%%)", n, por*100))
tbl_8 <- dcast(CELULAR~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_8, fill="0 (0.0%)")
write.xlsx(tbl_8, paste0(ruta,"tbl_8.xlsx"), sheetName = "tbl_8",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$CELULAR, Bd_Analisis$CONSUMOKwh_AGRUPADO))
```

```
tbl_9 <- Bd_Analisis %>%
  group_by(RADIO_COMUNICACION, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_9$export <- with(tbl_9, sprintf("%i (%.1f%%)", n, por*100))
tbl_9 <- dcast(RADIO_COMUNICACION~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_9, fill="0 (0.0%)")
write.xlsx(tbl_9, paste0(ruta,"tbl_9.xlsx"), sheetName = "tbl_9",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$RADIO_COMUNICACION, Bd_Analisis$CONSUMOKwh_AGRUPADO))
```

```
tbl_10 <- Bd_Analisis %>%
  group_by(INTERNET, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_10$export <- with(tbl_10, sprintf("%i (%.1f%%)", n, por*100))
tbl_10 <- dcast(INTERNET~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_10, fill="0 (0.0%)")
write.xlsx(tbl_10, paste0(ruta,"tbl_10.xlsx"), sheetName = "tbl_10",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$INTERNET, Bd_Analisis$CONSUMOKwh_AGRUPADO))
```

```
#####
#CARACTERÍSTICAS ELECTRICAS
#####
```

```
#ENERGIA_ELECTRICA
#MEDIDOR
#RECIBO
#FUENTE_ILUMINACION
#REFRIGERADOR
#AIRE_ACONDICIONADO
#CALEFACCION
```

```
tbl_11 <- Bd_Analisis %>%
  group_by(ENERGIA_ELECTRICA, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_11$export <- with(tbl_11, sprintf("%i (%.1f%%)", n, por*100))
tbl_11 <- dcast(ENERGIA_ELECTRICA~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_11, fill="0 (0.0%)")
write.xlsx(tbl_11, paste0(ruta,"tbl_11.xlsx"), sheetName = "tbl_11",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$ENERGIA_ELECTRICA, Bd_Analisis$CONSUMOKwh_AGRUPADO))
```

```
tbl_12 <- Bd_Analisis %>%
  group_by(MEDIDOR, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_12$export <- with(tbl_12, sprintf("%i (%.1f%%)", n, por*100))
```

```
tbl_12 <- dcast(MEDIDOR~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_12, fill="0 (0.0%)")
write.xlsx(tbl_12, paste0(ruta,"tbl_12.xlsx"), sheetName = "tbl_12",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$MEDIDOR, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_13 <- Bd_Analisis %>%
  group_by(RECIBO, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_13$export <- with(tbl_13, sprintf("%i (%.1f%%)", n, por*100))
tbl_13 <- dcast(RECIBO~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_13, fill="0 (0.0%)")
write.xlsx(tbl_13, paste0(ruta,"tbl_13.xlsx"), sheetName = "tbl_13",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$RECIBO, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_14 <- Bd_Analisis %>%
  group_by(FUENTE_ILUMINACION, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_14$export <- with(tbl_14, sprintf("%i (%.1f%%)", n, por*100))
tbl_14 <- dcast(FUENTE_ILUMINACION~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_14, fill="0 (0.0%)")
write.xlsx(tbl_14, paste0(ruta,"tbl_14.xlsx"), sheetName = "tbl_14",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$FUENTE_ILUMINACION, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_15 <- Bd_Analisis %>%
  group_by(REFRIGERADOR, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_15$export <- with(tbl_15, sprintf("%i (%.1f%%)", n, por*100))
tbl_15 <- dcast(REFRIGERADOR~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_15, fill="0 (0.0%)")
write.xlsx(tbl_15, paste0(ruta,"tbl_15.xlsx"), sheetName = "tbl_15",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$REFRIGERADOR, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_16 <- Bd_Analisis %>%
  group_by(AIRE_ACONDICIONADO, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_16$export <- with(tbl_16, sprintf("%i (%.1f%%)", n, por*100))
tbl_16 <- dcast(AIRE_ACONDICIONADO~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_16, fill="0 (0.0%)")
write.xlsx(tbl_16, paste0(ruta,"tbl_16.xlsx"), sheetName = "tbl_16",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$AIRE_ACONDICIONADO, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_17 <- Bd_Analisis %>%
  group_by(CALEFACCION, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_17$export <- with(tbl_17, sprintf("%i (%.1f%%)", n, por*100))
tbl_17 <- dcast(CALEFACCION~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_17, fill="0 (0.0%)")
write.xlsx(tbl_17, paste0(ruta,"tbl_17.xlsx"), sheetName = "tbl_17",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$CALEFACCION, Bd_Analisis$CONSUMOKwh_AGRUPADO))

#Características de como cocina
#LUGAR_COCINA
#COMBUSTIBLE_COCINAR
#LEÑA_Kg/DIA
#LEÑA_Kg_MES
#APARATOS_ELECTRICOS_Kwh_MES
```



```

tbl_18 <- Bd_Analisis %>%
  group_by(LUGAR_COCINA, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_18$export <- with(tbl_18, sprintf("%i (%.1f%%)", n, por*100))
tbl_18 <- dcast(LUGAR_COCINA~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_18, fill="0 (0.0%)")
write.xlsx(tbl_18, paste0(ruta,"tbl_18.xlsx"), sheetName = "tbl_18",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$LUGAR_COCINA, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_19 <- Bd_Analisis %>%
  group_by(COMBUSTIBLE_COCINAR, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )
tbl_19$export <- with(tbl_19, sprintf("%i (%.1f%%)", n, por*100))
tbl_19 <- dcast(COMBUSTIBLE_COCINAR~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_19, fill="0 (0.0%)")
write.xlsx(tbl_19, paste0(ruta,"tbl_19.xlsx"), sheetName = "tbl_19",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$COMBUSTIBLE_COCINAR, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_20 <- Bd_Analisis %>%
  group_by(LENA_Kg_MES, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_20$export <- with(tbl_20, sprintf("%i (%.1f%%)", n, por*100))
tbl_20 <- dcast(LENA_Kg_MES~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_20, fill="0 (0.0%)")
write.xlsx(tbl_20, paste0(ruta,"tbl_20.xlsx"), sheetName = "tbl_20",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$LENA_Kg_MES, Bd_Analisis$CONSUMOKwh_AGRUPADO))

tbl_21 <- Bd_Analisis %>%
  group_by(APARATOS_ELECTRICOS_Kwh_MES, CONSUMOKwh_AGRUPADO) %>%
  summarise(n = n()) %>%
  mutate(por = round(n/sum(n),1) )

tbl_21$export <- with(tbl_21, sprintf("%i (%.1f%%)", n, por*100))
tbl_21 <- dcast(APARATOS_ELECTRICOS_Kwh_MES~CONSUMOKwh_AGRUPADO, value.var="export", data = tbl_21, fill="0 (0.0%)")
write.xlsx(tbl_21, paste0(ruta,"tbl_21.xlsx"), sheetName = "tbl_21",
  col.names = TRUE, row.names = FALSE, append = FALSE)
chisq.test(table(Bd_Analisis$APARATOS_ELECTRICOS_Kwh_MES, Bd_Analisis$CONSUMOKwh_AGRUPADO))

#####
#####
#2. Modelación del consumo
#####
#####

#Recodificación del consumo es una variable cualitativa ordinal
table(Bd_Analisis$CONSUMOKwh_AGRUPADO)

Bd_Analisis$CONSUMOKwh_AGRUPADO[Bd_Analisis$CONSUMOKwh_AGRUPADO == '0 KWh'] <- '1'
Bd_Analisis$CONSUMOKwh_AGRUPADO[Bd_Analisis$CONSUMOKwh_AGRUPADO == '(0-40] KWh'] <- '2'
Bd_Analisis$CONSUMOKwh_AGRUPADO[Bd_Analisis$CONSUMOKwh_AGRUPADO == '(40-90] KWh'] <- '3'
Bd_Analisis$CONSUMOKwh_AGRUPADO[Bd_Analisis$CONSUMOKwh_AGRUPADO == '>90 KWh'] <- '4'

CONSUMO_DICO <- rep(1, dim(Bd_Analisis)[1])
Bd_Analisis <- cbind.data.frame(Bd_Analisis, "CONSUMO_DICO" = CONSUMO_DICO)
Bd_Analisis$CONSUMO_DICO[Bd_Analisis$CONSUMOKwh_AGRUPADO == '1'] <- 0

```

```

Bd_Analisis$CONSUMO_DICO[Bd_Analisis$CONSUMOKwh_AGRUPADO == '2'] <- 0
table(Bd_Analisis$CONSUMO_DICO)
table(Bd_Analisis$CONSUMOKwh_AGRUPADO)

#Modelo de regresión logística
Modelo <- glm(CONSUMO_DICO ~ UBICACION + USO + RADIO_COMUNICACION + ENERGIA_ELECTRICA + RECIBO +
FUENTE_ILUMINACION + REFRIGERADOR + LUGAR_COCINA + LENA_Kg_MES,
             family = binomial(link=logit),
             data = Bd_Analisis)

summary(Modelo)

Modelo_2 <- glm(CONSUMO_DICO ~ UBICACION + RADIO_COMUNICACION + RECIBO + REFRIGERADOR +
LENA_Kg_MES,
              family = binomial(link=logit),
              data = Bd_Analisis)

summary(Modelo_2)

```

## Anexo II. Análisis Correlacional

```

close all;
%% An-lisis de CorrelaciÛn
A = importdata('BASEJULIO2018.xls'); %Tarda cerca de un minuto
Data=A.data.BaseDeDatos;
BASE=[Data(:,11:19) Data(:,25:26) Data(:,36:38) Data(:,41:42) Data(:,45:48) Data(:,62) Data(:,64) Data(:,67) Data(:,70)
Data(:,73) Data(:,78) Data(:,87) Data(:,89:90) Data(:,108) Data(:,113) Data(:,115:118)];
%
save('BASEDATOSPERS.mat','A','BASE','Data');
load BASEDATOSPERS.mat
% VARIABLES EXTRAIDAS DE LA BASE ORIGINAL-> Archivo 'BASEJULIO2018.xls'
% Cu-ntos hogares residen en la vivienda?          columna 11 en la base de datos (columna K en excel)
% Cu-ntas personas componen el hogar
% Por favor indique cu-ntas salas tiene el hogar?
% Por favor indique cu-ntas cocinas tiene el hogar?      4
% Por favor indique cu-ntos comedores tiene el hogar?    5
% Por favor indique cu-ntos baÑos tiene el hogar?
% Por favor indique cuantas habitaciones tiene el hogar?
% Por favor indique cu-ntos garages tiene el hogar?      8
% Por favor indique cu-ntas bodegas tiene el hogar?

% Cu-l es el material predominante de las paredes exteriores?  10
% Cu-l es el material predominante de los pisos?              11

% La vivienda cuenta con telÈfono celular                    12
% La vivienda cuenta con equipo de radio para comunicaciones  13
% La vivienda cuenta con internet                             14

%Cu-ntos dÍas a la semana tiene el servicio de energÍa elÈctrica?  15
%Cu-ntas horas al dÍa tiene el servicio?                      16

% Valor pagado          17
% Consumo en kWh        18
% Estrato socioeconÛmico      19
% Promedio de los 'ltimos seis meses      20

% Consumo lamparas incandescentes Mcal Mes .  21

% Consumo lamparas ahorradores Mcal Mes .  22

% Consumo refrigeraciÛn Mcal Mes .  23

% Consumo ambiente Mcal mes      24

% Consumo calefacciÛn Mcal mes    26

% Consumo estufas electricas McalMes .  26

% Consumo gas Mcal mes    27

% Consumo carbÛn Mcal mes    (28)
% PodrÍa indicar la cantidad de leÑa que utiliza un dia para cocinar (Peso
% en Kgs) .  29

% Consumo de leÑa Mcal Mes .  30

% Consumo hornos Mcal mes    31

% Consumo aparatos electricos Mcal mes .  32
% Ingreso mensuales .  33

```

```

% Viento      34
% sol         35

%Para tener en cuenta el consumo de gas (col 26) es cero cuando aparece vacío
% Pero: Valor pagado (17), Consumo en kWh (18), Estrato socioeconómico (19)
% y Promedio de los últimos seis meses (20) no se pueden completar con ceros

%% Etiquetas 1.NumEncuesta 2.Latitud 3.Longitud 4.FactorExpSubregión 5.FactorExpMupio 6.Subregión 7.Municipio
% 8.Ubicación 9.Localidad
BASE_Etiq=[Data(:,1:9)];
BASE_Etiquetas=[];
%%
BASE1=[]; %Filtración por estrato col 19, se suprimen las filas donde no se encuentra información de estrato
for i=1:2479
    if not(isnan(BASE(i,19)))
        BASE1=[BASE1; BASE(i,:)];
        BASE_Etiquetas=[BASE_Etiquetas; BASE_Etiq(i,:)];
    end
end

% se ubican ceros en los datos vacíos o reconocidos como NaN

for i=1:size(BASE1,1)
    for j=1:35
        if isnan(BASE1(i,j))
            BASE1(i,j)=0;
        end
    end
end

% Si persisten algunos NaN, se quitan por filas

%% Procesa los datos para determinar la matriz de correlación
MC=corr(BASE1); % BASE1 es la base de datos depurada de la original BASE
[X,Y] = meshgrid(1:35,1:35);
figure;
tamano=get(0,'ScreenSize');
figure('position',[tamano(1) tamano(2) tamano(3) tamano(4)]);
surface(X,Y,MC);
title('Matriz de Correlación');

ax=gca; ax.XTick=[1:35]; ax.XTickLabelRotation=60;
ax.XTickLabel = {'Hogares
vivienda?','PersonasHogar','SalasHogar','CocinasHogar','ComedoresHogar','BañosHogar','HabitacionesHogar','GarajesHogar','Bo
degasHogar',...

'MaterialParedes','MaterialPisos','Teléfono','Radio','Internet','DíasEnergía','HorasEnergía','PagoEnergía','Consumo','Estratificaci
ón',...

'Promedio6Meses','McalIncandescentes','McalAhorradoras','McalRefrigeración','McalAmbiente','McalCalefacción','McalEstufa
E',...
'McalGas','McalCarbón','KgLeña','McalLeña','McalHornos','McalAparElectricos','Ingresos','Viento','Irradiancia'};

ay=gca; ay.YTick=[1:35]; ay.YTickLabelRotation=10;
ay.YTickLabel = {'Hogares
vivienda?','PersonasHogar','SalasHogar','CocinasHogar','ComedoresHogar','BañosHogar','HabitacionesHogar','GarajesHogar','Bo
degasHogar',...

```

```

'MaterialParedes','MaterialPisos','TelÈfono','Radio','Internet','DÌasEnergÌa','HorasEnergÌa','PagoEnergÌa','Consumo','Estratificaci
Ûn',...

'Promedio6Meses','McalIncandescentes','McalAhorradoras','McalRefrigeraciÛn','McalAmbiente','McalCalefacciÛn','McalEstufa
E',...
'McalGas','McalCarbÛn','KgLeÒa','McalLeÒa','McalHornos','McalAparElectricos','Ingresos','Viento','Irradiancia'};
e=ay.YTickLabel; %Etiqueta
disp('oprima enter'); pause;

etiquetas={'N´mero hogares por vivienda','N´mero personas por hogar','N´mero salas por hogar',...
'N´mero cocinas por hogar','N´mero comedores por hogar',...
'N´mero baÒos por hogar','N´mero habitaciones por hogar','N´mero garages por hogar',...
'N´mero bodegas por hogar','Material predominante paredes','Material predominante pisos',...
'Vivienda con telÈfono celular','Vivienda con radio comunicaciones','Vivienda con internet',...
'DÌas semana servicio energÌa','Horas dÌa servicio energÌa',...
'Valor pago energÌa elÈctrica mes','Consumo energÌa elÈctrica mes (kWh)','Estrato socioeconÛmico','Promedio consumo energÌa
'ltimos 6 meses (kWh)',...
'L·mparas incandescentes (Mcal mes)','L·mparas ahorradoras (Mcal mes)',...
'RefrigeraciÛn (Mcal mes)','Ambiente (Mcal mes)','CalefacciÛn (Mcal mes)','Estufas electricas (Mcal mes)',...
'Gas (Mcal Mes)','CarbÛn (Mcal mes)','LeÒa para cocinar dia (Kg)','Consumo de leÒa en (Mcal mes)',...
'Hornos (Mcal mes)','Aparatos elÈctricos (Mcal mes)','Ingresos mensuales','Recurso eÛlico','Recurso solar'};

%% Umbral
%Se crea el mapa de colores. MC representa la matriz de correlaciÛn
map = [[1 1 1]*1; [1 1 1]*0.975; [1 1 1]*0.95; [1 1 1]*0.925; [1 1 1]*0.9; [1 1 1]*0.875; [1 1 1]*0.85; [1 1 1]*0.825; [1 1
1]*0.8; [1 1 1]*0.775 ];
bw=im2bw(abs(MC),0.25); % Imagwen binaria 1ÿs y 0ÿs para un umbral de 0.1333
figure;
for i=2:size(MC,1)
    for j=1:i-1
        if bw(i,j)==1
            x=BASE1(:,i); y=BASE1(:,j);

            N=ndhist(x,y,'filter');
            imagesc(N); colormap(map);
            xlabel(etiquetas(i)); ylabel(etiquetas(j)); colorbar;
            titulo=[etiquetas{1,i} ' vs ' etiquetas{1,j}]; title(titulo);
            titulo=[titulo '.jpeg'];
            exportfig(gcf, titulo,'FontSize', 8, 'color', 'cmyk','Resolution',300, 'Width',6, 'Height', 5);
            %i    j
            disp('oprima enter'); %pause;
        end
    end
end
end
end

```

### Anexo III. Correspondencia entre variables asociadas

#### 1. Tablas de contingencia

Una tabla de contingencia es una de las formas más comunes de resumir datos categóricos. En general, el interés se centra en estudiar si existe alguna asociación entre dos variables cualitativas y/o calcular la intensidad de dicha asociación.

Sean  $X, Y$  dos variables categóricas con  $I$  y  $J$  categorías respectivamente. Un sujeto puede venir clasificado en una de las  $I \times J$  categorías, que es el número posible de categorías que existe.

##### 1.1.Distribución conjunta

Viene dada por

$$p_{ij} = P(X = i, Y = j) \quad \text{con } i = 1, \dots, I \text{ y } j = 1, \dots, J$$

Es la probabilidad de  $(X, Y)$  en la celda de la fila  $i$  de la columna  $j$ .

##### 1.2.Distribución marginal

$$p_{i+} = P(X = i) = \sum_{j=1}^J p_{ij}$$

$$p_{+j} = P(X = j) = \sum_{i=1}^I p_{ij}$$

##### 1.3.Distribución condicional

En la mayor parte de las tablas de contingencia, como en el ejemplo anterior, una de las variables, por ejemplo  $Y$ , es una variable de respuesta y la otra variable  $X$  es una variable explicativa o predictora. En esta situación no tiene sentido hablar de distribución conjunta, sino de distribución condicional:

$$P(Y = j | X = i) = p_{j|i} = \frac{p_{ij}}{p_{i+}}$$

##### 1.4.Independencia

Dos variables son independientes si:

$$p_{ij} = p_{i+} \cdot p_{+j}$$

es decir, que la distribución condicionada es igual a la marginal.

Si  $X$  y  $Y$  son variable respuesta entonces se habla de independencia. Si  $Y$  es variable respuesta y  $X$  es variable explicativa entonces se habla de homogeneidad.

##### 1.5.Estadístico de prueba para evaluar asociación

$$T = \sum_{i=1}^k \frac{(f_0 - f_e)^2}{f_e}$$

donde,

$f_0$  es la frecuencia observada.

$f_e$  es la frecuencia esperada.

Este estadístico de prueba se contrasta con el estadístico crítico  $\chi^2_{((c-1) \times (f-1))}$ , si existe asociación entre las variables evaluadas, el valor-p asociado a la prueba será menor que la significancia asumida.

## Anexo IV. Modelo de regresión logística, el método de estimación y la prueba de hipótesis

### Modelo Lineal Generalizado (MLG)

#### 1. Componentes del modelo:

- Componente aleatoria: Identificar la variable de respuesta ( $Y$ ) y especificar o asumir una distribución de probabilidad para ella.
- Componente sistemática: Especifica cuáles son las variables explicativas o predictoras. Estas variables entran de forma lineal.

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- Enlace: Especifica la relacional entre la media o el valor esperado del componente aleatorio (es decir,  $E(X)$ ) y el componente sistemático.

Para el caso del Modelo de Regresión Lineal:

$$Y = \beta_0 + \beta x_i + \epsilon_i$$

- Componente aleatoria:** Donde  $Y$  es la variable de respuesta, además  $\epsilon_i \sim N(0, \sigma^2)$ .
- Componente sistemática:** Donde  $X$  es la variable explicativa, donde se asume una relación lineal,

$$\beta_0 + \beta x_i$$

- Función de enlace:**

$$g(E[Y_i]) = E[Y_i] = \beta_0 + \beta x_i$$

Sea  $n$  el tamaño de muestra y supongamos que tenemos  $Y_1, Y_2, \dots, Y_n$  observaciones sobre nuestra variable de respuesta y que las observaciones son independientes entre ellas. Las  $Y$  son variables discretas donde  $Y$  puede ser:

Dicotómica (binaria): Con un número fijo de ensayos. Por ejemplo: éxito/falla; correcto/incorrecto; de acuerdo/en desacuerdo etc. Una solución sería asumir una respuesta Binomial.

Dado que, en una regresión lineal, se modelan los promedios. El enfoque está en el valor esperado de la variable de respuesta:

$$E(Y) = \mu$$

Ahora, si queremos investigar el comportamiento de la variación de  $\mu$  como una función de los niveles del predictor o variable explicativa  $X$ . La componente sistemática del modelo está compuesta de un conjunto de variables explicativas y alguna función lineal de ellos:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Esta combinación lineal de las variables explicativas se conoce como un predictor lineal.

Ahora bien, la función de enlace es la que se encarga de unir la componente aleatoria y la componente sistemáticas. La pregunta es cómo  $\mu = E(Y)$  se relaciona con  $\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ , esto se realiza a partir de la función de enlace  $g(\mu)$ , de esta manera:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Los enlaces siguientes son espaciales (según la distribución asumida):



Distribución	Parámetro natural	Enlace
Normal	$\mu$	Identidad
Poisson	$\log(\mu)$	Log
Binomial	$\log\left[\frac{\mu}{1-\mu}\right]$	logit

## 2. Distribuciones de la familia exponencial

$$f(y_i; \theta_i) = \exp\{y_i b(\theta_i) - c(\theta_i) + d(y_i)\}$$

Cuando se puede expresar la función de densidad de esta forma, entonces  $b(\theta_i)$  se denota como el parámetro natural.

## 3. Regresión logística

Modelar la probabilidad para datos binarios no es un problema de regresión lineal simple.

- La varianza de las respuestas dicotómicas  $Y$  para cada sujeto depende de  $X$ .
- La varianza no es constante a través de los valores de  $X$ , sino que es igual a  $V(Y) = n \cdot \pi(x) \cdot [1 - \pi(x)]$

Este modelo busca explicar la probabilidad de que se presente el evento en términos de la función logística:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

En el caso de la regresión lineal múltiple:

- $x_i = (1, x_{i1}, \dots, x_{ip-1})$ , un vector  $p$ -dimensional de predictores.
- $\beta = (\beta_0, \dots, \beta_{p-1})$  es un vector  $p$ -dimensional de coeficientes de regresión.

Ahora si expresamos el modelo de la siguiente manera:

$$\frac{\pi(x)}{1 - \pi(x)} = \exp\{\beta_0 + \beta_1 x\}$$

Si aplicamos el logaritmo:

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x$$

Esta expresión se conoce como el logit, ahora este si se requiere evaluar el cambio

de un evento cuando  $x$  cambia en una unidad:

$$\frac{\pi(x)/[1 - \pi(x+1)]}{\pi(x)/[1 - \pi(x)]} = \frac{\exp\{\beta_0 + \beta_1 x + \beta_1\}}{\exp\{\beta_0 + \beta_1 x\}} = \exp\{\beta_1\}$$

Esto indica que un aumento en  $x$  en unidad, la chance del evento aumenta en el factor  $\exp\{\beta_1\}$ , independiente del valor de  $x$ .

### 3.1. Método de estimación

Los coeficientes de regresión logística se estiman por el método de Máxima Verosimilitud. La función de verosimilitud es:

$$L(\beta_0, \beta_1) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

Entonces  $\hat{\beta}_0$  y  $\hat{\beta}_1$  maximizan  $L(\beta_0, \beta_1)$ .

### 3.2.Prueba de hipótesis

- La hipótesis asociada a cada variable explicativa:  $H_0: \beta_k = 0$  Vs  $H_1: \beta_k \neq 0$
- Estadístico de prueba:  $\frac{\hat{\beta}_1}{EE(\hat{\beta}_1)}$  si el valor-p asociado a cada variable explicativa es menor que la significancia ( $\alpha$ ) se dice que la variable aporta en la variación de la respuesta, por lo tanto es una variable importante en el modelo.

### Anexo IV. Conjunto de variables usadas para el análisis variables en rojo no se tuvieron en cuenta

Número	Número de encuesta
F.Expsubregión	Factor de expansión por subregión
F.Exmunicipio	Factor de expansión por municipio
Subregión	Subregión
Municipio	Municipio
Ubicación	Datos de la ubicación de la vivienda
Localidad	Nombre de la localidad
P.4	Tipo de vivienda
P.4.1	Cuál?
P.5	Cuántos hogares residen en la vivienda? (familiares o no)
P.6	Cuántas personas componen el hogar
P.7.1	Por favor indique cuántas salas tiene el hogar?
P.7.2	Por favor indique cuántas cocinas tiene el hogar?
P.7.3	Por favor indique cuántos comedores tiene el hogar?
P.7.4	Por favor indique cuántos baños tiene el hogar?
P.7.5	Por favor indique cuántas habitaciones tiene el hogar?
P.7.6	Por favor indique cuántos garages tiene el hogar?
P.7.7	Por favor indique cuántas bodegas tiene el hogar?
P.7.8	Otro, Cuál?
P.7.9	Número
P.8	La vivienda es
P.9	El uso de la vivienda es exclusivamente
P.9.4	Otro, Cuál?
P.10	Cuál es el material predominante de las paredes exteriores?
P.11	Cuál es el material predominante de los pisos?
P.12	De dónde proviene generalmente el agua para consumo humano?
P.13	Cuánto tiempo o cuánta distancia recorre para llevar el agua a su casa?
P.13.1	Unidad de tiempo
P.13.2	Tiempo recorrido para llevar el agua a su casa (Horas)
P.14	La vivienda cuenta con alcantarillado
P.15	El servicio sanitario es
P.16	De qué forma eliminan las basuras?
P.16.6	Otra, cuál?
P.17.1	La vivienda Cuenta con teléfono fijo con línea
P.17.2	La vivienda cuenta con teléfono celular
P.17.3	La vivienda cuenta con equipo de radio para comunicaciones
P.17.4	La vivienda cuenta con internet
P.18	La vivienda utiliza servicio de energía eléctrica?
P.19	Cuenta con medidor de energía eléctrica?
P.20	Cuántos días a la semana tiene el servicio de energía eléctrica?
P.21	Cuántas horas al día tiene el servicio?
P.22	Podría facilitarnos un recibo de pago?
P.23.1	Período facturado
P.23.2	Valor pagado
P.23.3	Consumo en kWh
P.23.4	Estrato socioeconómico
P.23.5	Promedio de los últimos seis meses
P.24.1	Podría decirnos cuánto pagó por el servicio el último mes o un mes que recuerde? MES
P.24.2	Valor pagado
P.25	En una semana normal con qué frecuencia tiene interrupciones del servicio de energía eléctrica?
P.26.1	Si hay interrupciones indique Horas
P.27	De contar con energía eléctrica de manera confiable usted estaría dispuesto a pagar una tarifa mensual?
P.28	Por qué razón no estaría dispuesto a pagar?

P.28.4	Otra, Cuál?
P.29	Por la conexión o acometida al servicio de la red usted estaría dispuesto a pagar:
P.30	Cuanto estaría dispuesto a pagar mensualmente, una vez que cuente con el servicio?
P.31	Cuál es la fuente de iluminación principal en la vivienda?
P.32.1	En promedio cuántas horas al día utiliza esta fuente de iluminación? Horas
P.32.2	Minutos
P.33.1	Consumo lamparas incandescentes kWh mes
P.33.1.1	Consumo lamparas incandescentes Mcal Mes
P.33.2	Consumo lamparas ahorradores kWh mes
P.33.2.1	Consumo lamparas ahorradores Mcal Mes
P.33.3	Consumo lamparas fluorescentes kWh mes
P.33.3.1	Consumo lamparas fluorescentes Mcal Mes
P.34	Usa nevera o refrigerador?
P.35	Consumo refrigeración kWh mes
P.35.1	Consumo refrigeración Mcal Mes
P.36	Usa aire acondicionado o ventilador?
P.37	Consumo ambiente kWh Mes
P.37.1	Consumo ambiente Mcal mes
P.38	Usa algún sistema para agua caliente o calefacción
P.39	Consumo calefacción kWh mes
P.39.1	Consumo calefacción Mcal mes
P.40	En qué lugar de la vivienda cocinan
P.41	En general qué combustible usa principalmente para cocinar?
P.41.8	Otro, cuál?
P.42	Consumo estufas electricas kWhMes
P.42.1	Consumo estufas eléctricas McalMes
P.44	Utiliza una segunda estufa para cocinar?
P.45	Cada cuánto usa esta segunda estufa?
P.45.2	Unidad de medida
P.45.3	Frecuencia de uso al mes de la segunda estufa electrica
P.46	Podría indicar que combustible utiliza para la segunda estufa?
P.46.8	Otro, cuál?
P.47	Consumo de gas galones Mes
P.47.1	Consumo gas kWh mes
P.47.1.1	Consumo gas Mcal mes
P.47.2	Consumo carbón Kg/Mes
P.47.2.1.1	Consumo carbón Mcal mes
P.48	Podría indicar la cantidad de leña que utiliza un día para cocinar (Peso en Kgs)
P.48.2	A qué especie de madera pertenece la leña?
P.49	De dónde extrae la leña para cocinar?
P.50	Cada cuánto recolecta leña para cocinar? Cantidad
P.50.1	Frecuencia
P.50.2	Frecuencia de recolección al mes
P.49.5	Otro, cuál?
P.51.1	Cuánto tiempo o cuánta distancia recorre para llevar su leña a su casa? Tiempo
P.51.2	Unidad de tiempo
P.51.3	Tiempo que recorre para llevar la leña a la casa (Horas)
P.52	Dónde compra regularmente la leña?
P.52.5	Otra, cuál?
P.53	Qué tipo de estufa de leña tiene?
P.53.5	Otra, cuál?
P.54.1	A qué hora prende la estufa? 24 horas
P.54.3	La mantiene prendida todo el día?
P.47.3	Consumo de leña Kg Mes
P.47.3.1	Consumo de leña kWh Mes
P.47.3.2	Consumo de leña Mcal Mes
P.55	Utiliza el horno para preparar alimentos?
P.56	Qué combustible utiliza para el horno?

P.56.7	Otro, cuál?
P.57	Consumo hornos kWh mes
P.57.1	Consumo hornos Mcal mes
P.59	Consumo aparatos electricos kWh mes
P.59.1	Consumo aparatos electricos Mcal mes
P.62	Ingreso mensuales